Molecular Studies of Saline Antarctic Lakes from a Whole Ecosystem Perspective

Sheree Yau

2012

A thesis submitted for the degree of Doctor of Philosophy

School of Biotechnology and Biomolecular Sciences
University of New South Wales, Australia

Originality Statement

'I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published of written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.'

Signed ....

Date ....

Copyright Statement

Authenticity Statement

'I certify that the Library deposit digital copy is a direct equivalent of the final officially approved version of my thesis. No emendation of content has occurred and if there are any minor variations in formatting, they are the result of the conversion to digital format.'

Acknowledgements

List of Publications

Publications and submitted manuscripts arising from my PhD research are listed below. In all cases, my supervisor Prof Ricardo Cavicchioli and my co-supervisor Dr Federico Lauro were involved in the research design and editing of the manuscripts. Where versions of published material appears in this thesis, details of the contributions made by myself and others precede it.

- David Wilkins, **Sheree Yau**, Timothy Williams, Michelle Allen, Mark V. Brown, Matthew Z. DeMaere, Federico M. Lauro and Ricardo Cavicchioli. Key Microbial Drivers in Antarctic Aquatic Environments. *FEMS Microbiology Reviews* (in press), 2012.

- **Sheree Yau** and Ricardo Cavicchioli. Microbial communities in Antarctic lakes: Entirely new perspectives from metagenomics and metaproteomics. *Microbiology Australia* 32:157–159, 2011.

- Federico M. Lauro, Matthew Z. DeMaere, **Sheree Yau**, Mark V. Brown, Charmaine Ng, David Wilkins, Mark J. Raftery, John A.E. Gibson, Cynthia Andrews-Pfannkoch, Matthew Lewis, Jeffery M. Hoffman,Torsten Thomas and Ricardo Cavicchioli. An integrative study of a meromictic lake ecosystem in Antarctica. *ISME Journal* 5:879–895, 2011.

- **Sheree Yau**, Federico M. Lauro, Matthew Z. DeMaere, Mark V. Brown, Torsten Thomas, Mark J. Raftery, Cynthia Andrews-Pfannkoch, Matthew Lewis, Jeffery M. Hoffman, John A. Gibson and Ricardo Cavicchioli. Virophage control of antarctic algal host–virus dynamics. *Proceedings of the National Academy of Sciences USA* 108:6163–6168, 2011.

- **Sheree Yau**, Timothy J. Williams, Federico M. Lauro, Matthew Z. DeMaere, Mark V. Brown, John Rich, John A.E. Gibson, Ricardo Cavicchioli. Heterotrophic resourcefulness and unusual sulfur biogeochemistry in a hypersaline lake. <u>*ISME Journal*</u> submitted, 2013.

Abstract

Contents

List of Figures

List of Tables

List of Abbreviations

1D-SDS PAGE one dimensional-sodium dodecyl sulphate polyacrylamide gel electrophoresis

AAnP aerobic anoxygenic photosynthesis

ABC ATP-binding cassette

ANOSIM analysis of similarity

BchlA bacteriochlorophyll A

BLAST basic local alignment search tool

CAMERA Community Cyberinfrastructure for Advanced Microbial Ecology Research and Analysis

CAS CRISPR-associated proteins

COG clusters of orthologous groups

CRISPR clustered regularly interspaced short palindromic repeat

DGGE denaturing gradient gel electrophoresis

DMS dimethylsulphide

DMSO dimethylsulphoxide

DMSP dimethylsulphoprorionate

DNA deoxyribonucleic acid

DO dissolved oxygen

DOC dissolved organic carbon

DRP dissolved reactive phosphorus

GOS global ocean sampling

GSB green sulphur bacteria

HMMER biosequence analysis using profile hidden Markov models

JCVI J.Craig Venter Institute

KEGG Kyoto Encyclopedia of Genes and Genomes

KO KEGG Orthology

KOBAS KEGG Orthology Based Annotation System

MS mass spectrometry

MS-MS two dimensional mass spectrometry

NCBI National Center for Biotechnology Information

NR non-redundant database

ORF open-reading frame

OTU operational taxonomic unit

PCA Principal Component Analysis

PCR polymerase chain reaction

PCTE polycarbonate Track Etch^{TM}

PR proteorhodopsin

QIIME Quanititative Insights Into Microbial Ecology

RDP Ribosomal Database Project

RNA ribonucleic acid

rRNA ribosomal RNA

rTCA reverse tricarboxylic acid

RuBisCO ribulose-bisphosphate carboxylase oxygenase

SO Southern Ocean

SRB sulphate-reducing bacteria

SSU small subunit ribosomal RNA

STAMP Statistical Analysis of Metagenomic Profiles

TDN total dissolved nitrogen

TDP total dissolved phosphorus

TDS total dissolved sulphur

TIGRFAM the Institute of Genomic Research curated protein database

TN total nitrogen

TOC total organic carbon

TP total phosphorus

TS total sulphur

VLP virus-like particle

WGS whole genome shotgun

WL Wood-Ljungdahl; or reductive acetyl-CoA

Chapter 1

General introduction: molecular microbial ecology of Antarctic lakes

Co-authorship statement

Sections of this chapter has been published as:

I contributed the section of the publication entitled *Antarctic lakes* excluding the subsection, *Microbial mats as microcosms of Antarctic life*. This material appears in section 1.6 *Molecular approaches used in Antarctic lake systems*, section 1.7 *Molecular insights into Antarctic lakes* and section 1.8 *limitations of taxonomic surveys* of this introduction.

1.1 Introduction

Antarctica is a "frozen desert" of constant low temperature, little precipitation and long polar light cycles where only specially adapted organisms can survive. The continent is covered by ice up to 4 km thick that covers an area of 13.8 million km^2. A tiny 0.32% of the land area is ice-free, most of which consists of exposed rocky peaks or nunataks such as in the Ellsworth, the Transantarctic and the North Victoria Land Mountains. Only 1–2% of that ice-free land comprises rocky coastal oases; however, it is these regions where Antarctic life is concentrated (Hodgson, 2012).

They are breeding sites for large animals such as seals, penguins and sea birds and some of the only locations where plants and lichens are found. Coastal oases are also distinguished by the presence of hundreds of lakes and ponds. Life in these lakes is microbially dominated with few, if any, metazoan inhabitants (Laybourn-Parry, 1997) making them ideal locations to study Antarctic microbiota. The lakes span a continuum of environmental factors such as salinity and are "natural laboratories" to examine adaptations to a property of interest. They are also ideal model ecosystems as they are largely isolated with a close relation between species and function.

This introduction will describe the Antarctic lakes, their microbiology and review molecular-based Antarctic microbiological research on the lakes. As this thesis focused on two lakes in the Vestfold Hills, emphasis will be given to describing research from this study site.

1.2 Antarctic lakes

In Antarctica, perenially available liquid water is found predominantly in lakes. These lakes span a wide range of physical and chemical properties from freshwater to hypersaline and constantly ice-covered to melted. Some are permanently stratified and termed meromictic if they thaw seasonally, or amictic if they are always ice-covered. Stratified lakes provide a unique opportunity to describe microbial populations along chemical gradients, but within a single water body. Most lakes are ice-covered for most of the year making them effectively isolated, and some may be truly closed systems if ice-cover is permanent. The age of water varies considerably; for example, outflow of subglacial water at Blood Falls is estimated to be 1.5 million years

old (Mikucki *et al.*, 2009) while water from Lake Miers is less than 300 years old (Green *et al.*, 1988). Overall, there are two main Antarctic lake types: those bound by ice, comprising subglacial; epiglacial and supraglacial lakes, and those bound by rock.

1.2.1 Ice-bound lakes

Subglacial lakes are pools of water beneath an ice sheet that form as the pressure of glacier flow against bedrock melts the ice at the interface. They are prevalent in Antarctica with at least 145 identified (Siegert *et al.*, 2005). The largest of these is Lake Vostok, which is 240 km long, 50 km wide and up to 1 km deep (Siegert *et al.*, 2001). Here the pressure is an extreme 340 atmospheres (Siegert *et al.*, 2001). These lakes are found dotted around the continent generally under the continental ice shelf (Siegert *et al.*, 2001).

Epiglacial lakes are similarly formed in the boundary between rock and ice, but where rock is exposed, such as where a glacier front contacts a mountain side. They are potentially the most common lake type as they can occur where ever there is rock and thus are both inland and coastal (Hodgson, 2012). These lakes can be highly changeable as they are subject to glacial movements and meltwater inputs. As a result, they can be short-lived, but many examples are thousands of years old, such as the .

Pockets of water can be also be found on top of glaciers. One example is cryoconite holes, which originate from the heat absorbed by dark dust melting small depressions on the glacier surface. These are extremely interesting systems with massive ranges in pH and chemistry. Most substantial supraglacial lakes also occur. However, all of the supraglacial reservoirs of water tend to be ephemeral lasting only during the summer months.

1.2.2 Rock-bound lakes

These lakes are of water trapped in exposed rocky basins. Look up the Mountain lakes.

By far the majority of lakes are found in the coastal oases. Most of these lakes were formed when the retreat of the continental ice-shelf lead to isostatic uplift of the land (Burton, 1981). As a result, the majority of lakes in the coastal oases are composed of relic seawater and are predominantly saline or hypersaline (Burke and Burton, 1988). In the latter, salinity is high due

to concentrated by ablation (evaporation and sublimnation). Lakes closer to the coastline may still occasionally experience marine inputs. Epishelf lakes are a type of lake unique to coastal regions.

Freshwater lakes near the continental ice shelf were likely already above sea-level as the ice receded and are not of marine origin (Bronge, 2004). Other freshwater lakes were originally marine-derived but have been flushed fresh by glacial meltwater (Pickard *et al.*, 1986). All lakes may receive water inputs from precipitation, from the ice-shelf and glacial melt streams (Burton, 1981). This can cause freshwater to seasonaly overlay some saline lakes as the ice-cover thaws. The chemistry of the exposed lakes is very much influenced by the water balance from local geographic and climatic conditions which leads them to have different physical and chemical properties.

1.3   Coastal oases

These are also the best studied systems as research stations are the most hospitable sites for research stations. Coastal oases, where lakes are found fringe the Antarctic continent. In East Antarctica these include the Vestfold Hills, Bunger Hills, Larsemann Hills, Syowa Oasis, Schirmacher Oasis, Grearson Hills and McMurdo Dry Valleys. In West Antarctic, the Peninsula, the sub-antarctic islands and maritime islands house multiple lakes. Of these locations, the best studied lake systems are those of the McMurdo Dry Valleys, The Vestfold Hills and the subantarctic islands.

1.4   The Vestfold Hills, East Antarctica

The Vestfold Hills is a ice-free region of approximately 400 km^2 on the eastern shore of the Prydz Bay, East Antarctica in the Australian Antarctic Territory (fig:vestfold map) (Gibson, 1999). The region is made up of three large peninsulae, Broad, Mule and Long Peninsula, separated by Fjords connected to the sea. Some of these are large, such as Ellis Fjord which is 10 km long, up to 100 m deep and has become a stratified system due to its restricted opening to the ocean (Burke and Burton, 1988). The region was formed approximately 10,000 years ago in the early Holocene from isostatic rebound (Burton, 1981).

The Vestfold Hills were first sighted and named in 1935 (Law, 1959).

Only intermittent expeditions occurred in the area until the establishment of Davis Station (68°33′S, 78°15′E) in 1957 (Law, 1959). A continuous presence has been maintained since. There Vestfold Hills region was immediately noted for its extensize ice-free land and the numerous lakes (Johnstone *et al.*, 1973).

The Australian Antarctic Data Centre lists more than 3,000 water bodies mapped in the Vestfold Hills, ranging in area from 1 to 8,757,944 m^2. More than 300 lakes and ponds have been described, including approximately 20% of the world's meromictic lakes (Gibson, 1999). These are of particular interest because the anoxic bottom waters help preserve a paleogeological record in the sediments. This can tell us about the region and particularly climatic changes. Stratified lakes provide a unique opportunity to describe microbial populations along chemical gradients, but within a single water body.

1.4.1   Biology of the Vestfold Hills

1.4.2   Lakes of the Vestfold Hills

1.5   Cultivation and microscopy-based Antarctic microbiology

Bacteria were detected by cultivation or by microscopy. Identification was limited to those species that could be isolated and appropriate identification tests performed. Eucarya were identified with microscopy based approaches.

1.5.1   Eucarya

1.5.2   Bacteria

1.5.3   Archaea

1.5.4   Viruses

As obligate parasites, culturing viruses is made problematic by the need to have a susceptible host in culture. Furthermore, host specificity can be extremely narrow so any assessment of viral diversity by cultivation is highly limited. This is compounded by the logistical constraints of conducting field work in the Antarctic.

Most studies of Antarctic viruses have been confined to diversity analyses based on the electron micrographs of virus-like particle (VLP) morphotypes or visibly infected cells. Electron micrographs are able to distinguish to some extent tailed bacteriophages such as myo- sipho- and podo- viruses from one another due to tail morphology. For tailess viruses with icosohedral symmetry, morphology alone provides hardly any distinguishing features apart from capsid size. Other metrics used to assess viruses in the environment include enumeration, calculated the virus to bacteria ratios, visibly infected cells and viral production rates.

Overall, pioneering work on viruses has made several noteworthy observations.

1. Viruses are possibly more abundant in high latitude lakes than lower latitude.

2. Burst sizes of viruses are lower in high latitude lakes than in low latitude.

3. Viral abundance appears to correlate positively with salinity.

4. As food chains in Antarctic Lakes are truncated, viruses may play an increased importance in Antarctic lakes, particularly in increasing secondary production through the microbial loop.

5.

6.

In the absence of metazoan grazers, viruses are hypothesized to play an increased role in the microbial loop in Antarctic systems (Kepner *et al.*, 1998) and as drivers of microbial evolution (Anesio and Bellas, 2011). This idea has been supported by microscopy-based observations of viral density, virus to bacteria ratios and infection rates that are different in Antarctic lakes than lower latitude systems (Laybourn-Parry *et al.*, 2001; Madan *et al.*, 2005; Laybourn-Parry and Pearce, 2007; Säwström *et al.*, 2007). Cold environments are hypothesized to be a 'hotspot' of viral diversity. However, molecular methods are required to validate this claim.

1.6 Molecular approaches used in Antarctic lake systems

The majority of molecular-based studies of Antarctic aquatic microbial communities have made use of polymerase chain reaction (PCR) amplification of small subunit ribosomal RNA (SSU) sequences to survey the diversity of *Bacteria* and in some cases *Archaea* and *Eucarya*. Microbial composition has been determined by cloning and sequencing of ribosomal RNA (rRNA) gene amplicons exclusively (Bowman *et al.*, 2000b,a; Gordon *et al.*, 2000; Christner *et al.*, 2001; Purdy *et al.*, 2003; Karr *et al.*, 2006; Matsuzaki *et al.*, 2006; Kurosawa *et al.*, 2010; Bielewicz *et al.*, 2011), although most studies have also made use of denaturing gradient gel electrophoresis (DGGE) to provide a molecular "fingerprint" of the community (Pearce, 2003; Pearce *et al.*, 2003; Karr *et al.*, 2005; Pearce, 2005; Pearce *et al.*, 2005; Unrein *et al.*, 2005; Glatz *et al.*, 2006; Mikucki and Priscu, 2007; Mosier *et al.*, 2007; Schiaffino *et al.*, 2009; Villaescusa *et al.*, 2010). Functional genes have also been targeted using polymerase chain reaction (PCR) amplification to assess the potential of biochemical processes occurring, such as nitrogen fixation (Olson *et al.*, 1998), ammonia oxidation (Voytek *et al.*, 1999), anoxygenic photosynthesis (Karr *et al.*, 2003), and dissimilatory sulfite reduction (Karr *et al.*, 2005; Mikucki *et al.*, 2009).

1.7 Insights from Antarctic molecular studies

1.7.1 Bacterial diversity: adaptation to unique physical and chemical conditions

The vast majority of molecular studies of Antarctic lakes have focused on bacteria. Consistent with the wide range of physical and chemical properties of Antarctic lakes, a large variation in species assemblages have been found. While exchange of microorganisms must be able to occur between lakes that are in close vicinity to each other, the picture that has emerged from the data to date is that microbial populations are relatively unique to each type of isolated system. Nonetheless, certain trends in bacterial composition are also apparent.

Focusing on the similarities, lakes of equivalent salinities tend to have similar communities. Hypersaline lakes from the Vestfold Hills (Bowman

et al., 2000a) and McMurdo Dry Valleys (Glatz *et al.*, 2006; Mosier *et al.*, 2007) were all dominated by *Gammaproteobacteria* and members of the Bacteroidetes as well as harboring lower abundance populations of *Alphaproteobacteria*, *Actinobacteria*, and *Firmicutes*. The surface waters of saline lakes resemble marine communities dominated by *Bacteroidetes*, *Alphaproteobacteria* and *Gammaproteobacteria*, but divisions such as Actinobacteria and specific clades of *Cyanobacteria* have been found to be overrepresented compared to the ocean (Lauro *et al.*, 2011). Sediments from saline lakes in the Vestfold Hills (Bowman *et al.*, 2000b) and Nuramake-Ike in the Syowa Oasis (Kurosawa *et al.*, 2010) were very similar, containing in addition to the surface clades, *Deltaproteobacteria*, *Planctomycetes*, *Spirochaetes*, *Chloroflexi* (green non-sulphur bacteria), *Verrucomicrobia* and representatives of candidate divisions. Plankton from freshwater lakes were characterized by an abundance of *Betaproteobacteria*, although *Actinobacteria*, *Bacteroidetes*, *Alphaproteobacteria* and *Cyanobacteria* were also prominent (Pearce, 2003, 2005; Pearce *et al.*, 2005; Schiaffino *et al.*, 2009).

Bacterial diversity defined by nutrients

Differences in bacterial community structure are also influenced by nutrient availability. In studies of freshwater lakes in the Antarctic Peninsula and the South Shetland Islands, cluster analysis of DGGE profiles grouped together lakes of similar trophic status (Schiaffino *et al.*, 2009; Villaescusa *et al.*, 2010). Most of the variance in community structure could be explained by related chemical parameters such as phosphate and dissolved inorganic nitrogen. Similarly, three freshwater lakes, Moss, Sombre and Heywood on Signy Island are alike except that Heywood Lake is enriched by organic inputs from seals.

Bacterial composition in each lake changed from winter to summer and this was again correlated to variation in physico-chemical properties (Pearce, 2005). The bacterial population of Heywood Lake had shifted from a dominance of *Cyanobacteria* towards a greater abundance of *Actinobacteria* and marine *Alphaproteobacteria* (Pearce *et al.*, 2005). This hints at a link between a copiotrophic lifestyle in the Heywood Lake *Actinobacteria* and inhibition of Antarctic freshwater *Cyanobacteria* by eutrophication. This type of study exemplifies how inferences can be made about taxa and function by examining population changes over time and over gradients of environmental

parameters.

Bacterial biogeography

The relative isolation and diverse chemistries of the lakes facilitates biogeographical and biogeochemical studies. The anoxic and sulfidic bottom waters of some meromictic lakes form due to a density gradient that precludes mixing. Although sedimentation from the upper aerobic waters may occur, there is little opportunity for interchange of species with the bottom water of lakes allowing for greater divergence in community composition as nutrients can become depleted and products of metabolism can accumulate. As a result, distinct distributions of bacterial groups can inhabit these strata, and different types of microorganisms can be found in equivalent strata in different lakes. A good example of this is the presence of common types of purple sulphur bacteria (*Chromatiales*)and green sulphur bacteria (GSB)(*Chlorobi*) in some meromictic lakes and stratified fjords in the Vestfold Hills (Burke and Burton, 1988), compared to diverse purple non-sulphur bacteria in Lake Fryxell in Victoria Land (Karr *et al.*, 2003). In Lake Bonney, the east and west lobes harbor overlapping but distinct communities in the suboxic waters (Glatz *et al.*, 2006). The east lobe was dominated by *Gammaproteobacteria* and the west lobe by *Bacteroidetes*, illustrating how divergent communities can form from the same seed population. In contrast, ice communities are more readily dispersed by wind, aerosols and melt-water. 16S rRNA gene probes designed from bacteria trapped in the permanent ice-cover of Lake Bonney hybridized to microbial mat libraries sourced up to 15 km away (Gordon *et al.*, 2000). This demonstrates how a single lake may encompass microorganisms that are geographically dispersed, while also harboring others that have restricted niches and are under stronger selection pressure.

Bacterial diversity of Lake Vostok

Subglacial systems, such as Lake Vostok, have been isolated from the open environment for hundreds of thousands to millions of years (Siegert *et al.*, 2001). As a result they provide a reservoir of microorganisms that may have undergone significant evolutionary divergence from the same seed populations that were not isolated by the Antarctic ice cover. Lake Vostok is approximately 4 km below the continental ice-sheet making it extremely

9

difficult to determine suitable means for accessing the lake without inadvertently contaminating it with biological or chemical matter (**?????**). To date, molecular microbial studies have concentrated on the accretion ice above the ice-water interface (Priscu, 1999; Christner *et al.*, 2001). Accretion ice has been found to contain a low density of bacterial cells from *Alphaproteobacteria*, *Betaproteobacteria*, *Actinobacteria* and *Bacteroidetes* divisions closely allied to other cold environments. Molecular signatures of a thermophilic *Hydrogenophilus* species were also identified in accretion ice raising the possibility that chemoautotrophic thermophiles were delivered to the accretion ice from hydrothermal areas in the lakes bedrock (Bulat *et al.*, 2004; **?**). However, interpretation of results from samples sourced from the Lake Vostok bore hole are very challenging as it is difficult to differentiate contaminants from native Vostok microorganisms. From a study that assessed possible contaminants present in hydrocarbon-based drilling fluid retrieved from the Vostok ice core bore hole, six phylotypes were designated as new contaminants (Alekhina *et al.*, 2007). Two of these were *Sphingomonas* phylotypes essentially identical to those found in the accretion ice-core (Christner *et al.*, 2001), which raises question about whether bacterial signatures identified from the ice-cores are representative of Lake Vostok water, and further highlights the ongoing problem of causing forward contamination into the lake.

1.7.2 *Archaea*: methanogens and haloarchaea

Archaea have been detected mainly in anoxic sediments and bottom waters from lakes that range in salinity from fresh to hypersaline, and those with known isolates are affiliated with methanogens or haloarchaea (Bowman *et al.*, 2000a,b; Purdy *et al.*, 2003; Kurosawa *et al.*, 2010; Lauro *et al.*, 2011). Anoxia allows for the growth of methanogenic *Archaea* that mineralize fermentation products such as acetate, H_2 and CO_2 into methane, thereby performing an important step in carbon cycling. The acetoclastic methanogens thrive in environments where alternative terminal electron acceptors such as sulfate and nitrate have been depleted.

One example of this is Lake Heywood where methanogenic *Archaea* were found to comprise 34% of the total microbial population in the freshwater sediment, the majority of which were *Methanosarcinales* which include acetate and C1-compound utilizing methanogens (Purdy *et al.*, 2003).

In general, archaeal populations appear to be adapted to their specific lake environment. Sediments from saline lakes of the Vestfold Hills were inhabited by members of the *Euryarchaeota* typically found in sediment and marine environments with the phylotypes differing between the lakes examined (Bowman *et al.*, 2000b). While a phylotype similar to *Methanosarcina* was identified, the majority were highly divergent. Similarly, *Methanosarcina* and *Methanoculleus* were detected in Lake Fryxell but other members of the *Euryarchaeota* and *Crenarchaeota* (a single sequence) were divergent, clustering only with marine clones (Karr *et al.*, 2006). Based on the lake chemical gradients and the location of these novel phylotypes in the water column the authors speculated these archaea may be have alternative metabolisms such as anoxic methanotrophy or sulphur-utilization.

In sediments from Lake Nurume-Ike in the Langhovde region, 205 archaeal clones grouped into three phylotypes, with the predominant archaeal clone being related to a clone from Burton Lake in the Vestfold Hills, while the other two did not match to any cultivated species (Kurosawa *et al.*, 2010). In hypersaline lakes where bottom waters do not become completely anoxic, methanogens are not present and *Archaea* have extremely low abundance. For example, only two archeael clones of the same phylotype were recovered from deep water samples from Lake Bonney (Glatz *et al.*, 2006), and Organic Lake in the Vestfold Hills had an extremely low abundance of archaeal clones related to *Halobacteriales* (Bowman *et al.*, 2000a). In contrast to these stratified hypersaline lakes, the microbial community in the extremely hypersaline Deep Lake is dominated by haloarchaea (Bowman *et al.*, 2000a). Many of the clones identified from Deep Lake are similar to *Halorubrum* (formerly *Halobacterium*) *lacusprofundi* which was isolated from the lake (Franzmann *et al.*, 1988).

1.7.3   *Eucarya* perform multiple ecosystem roles

Single-celled *Eucarya* are important members of Antarctic aquatic microbial communities. In many Antarctic systems, eucaryal algae are the main photosynthetic organisms and in others, only heterotrophic protists occupy the top trophic level. As eucaryal cells are generally large with characteristic morphologies, microscopic identifications have been used. However, microscopy is unable to classify smaller cells such as nanoflagellates with high resolution, although these may constitute a high proportion of algal biomass. For ex-

ample, five morphotypes of *Chrysophyceae*, evident in Antarctic lakes were unidentifiable by light microscopy but were able to be classified using DGGE and DNA sequencing (Unrein *et al.*, 2005). Consistent with this, molecular studies specifically targeting eucaryal diversity (Unrein *et al.*, 2005; Mosier *et al.*, 2007; Bielewicz *et al.*, 2011) have identified a much higher level of diversity than previously suspected, and the studies have discovered lineages not previously known to be present such as silicoflagellates (Unrein *et al.*, 2005) and fungi (Mosier *et al.*, 2007; Bielewicz *et al.*, 2011).

Most eucarya in Antarctic lakes are photosynthetic microalgae that are present in marine environments with a wide distribution including chlorophytes, haptophytes, cryptophytes and bacillariophytes. Molecular methods have afforded deeper insight into the phylogenetic diversity within these broader divisions and have revealed some patterns in their distribution. Using 18S rRNA gene amplification and DGGE, the same chrysophyte phylotypes were identified in lakes from the Antarctic Peninsula and King George Island despite being 220 km apart (Unrein *et al.*, 2005) indicating these species may be well-adapted to Antarctica or highly dispersed. Similarly, an unknown stramenopile sequence was detected throughout the 18S rRNA clone libraries of Lake Bonney demonstrating a previously unrecognized taxon occupied the entire photic zone in the lake (Bielewicz *et al.*, 2011). In contrast, other groups showed distinct vertical and temporal distributions with cryptophytes dominating the surface, haptophytes the midwaters and chlorophytes the deeper layers during the summer while stramenopiles increased in the winter (Bielewicz *et al.*, 2011). Further studies are necessary to determine the basis for apparent specific adaptations of some species to particular lakes or lake strata, and for the cosmopolitan distribution of others. Here, molecular based research of the kind that has been applied to bacteria such as functional gene surveys will undoubtedly help answer these questions.

1.7.4   Functional gene studies of Antarctic lakes

1.7.5   Integrative studies to derive whole ecosystem function

1.8   Limitations of taxonomic surveys

Inferring functional potential from taxonomic surveys can be problematic due to species or strain level differences in otherwise related bacteria. For example, the majority of the *Gammaproteobacteria* in hypersaline lakes were relatives of *Marinobacter* suggesting that this genus is particularly adapted to hypersaline systems (Bowman *et al.*, 2000b; Glatz *et al.*, 2006; Matsuzaki *et al.*, 2006; Mosier *et al.*, 2007). Nonetheless, *Marinobacter* species from different lakes appeared biochemically distinct as isolates from hypersaline lake Suribati-Ike were all able to respire dimethylsulphoxide (DMSO) but not nitrate (Matsuzaki *et al.*, 2006). In contrast, those from the west lobe of Lake Bonney were all able to respire nitrate (Ward and Priscu, 1997). Interestingly, in the east lobe of the same lake, nitrate respiration was inhibited although a near-identical *Marinobacter* phylotype was present; it was speculated that the inhibition may have been caused by an as yet unidentified chemical factor (Ward *et al.*, 2005; Glatz *et al.*, 2006).

This also applies to *Eucarya*, as the influence of flagellates on ecosystem function is not necessarily clear-cut as they can simultaneously inhabit several trophic levels. For instance, in Ace Lake the mixotrophic phytoflagellate *Pyramimonas gelidocola* derives a proportion of its carbon intake through bacterivory (Bell and Laybourn-Parry, 2003) but in the nearby Highway Lake, it uptakes dissolved organic carbon (Laybourn-Parry *et al.*, 2005). This again illustrates potential limitations for deriving ecosystem level functions from taxonomic studies alone, even with taxa that appear physiologically straightforward.

1.9   '-omics' approaches

Metagenomic studies have assessed both the taxonomic composition and genetic potential of lake communities, and in some cases have linked function to specific members of the community (López-Bueno *et al.*, 2009; Ng *et al.*, 2010; Lauro *et al.*, 2011; Yau *et al.*, 2011; Varin *et al.*, 2012). When coupled with functional "omic" techniques (to date metaproteomics has been ap-

plied, but not metatranscriptomics or stable isotope probing), information has also been gained about the genetic complement that has been expressed by the resident populations (Ng *et al.*, 2010; Lauro *et al.*, 2011). More detailled description of the advances of '-omics' approaches is detailed in the introduction of chapter 2.

Metagenomics has enabled unprecedented insight into viral diversity and ecology by permitting more precise classification, information on genetic content and discovery of novel species. Metagenomic analysis of the virione of the freshwater Lake Limnopolar, Livingston Island uncovered the greatest depth of viral diversity of any aquatic system to date (López-Bueno *et al.*, 2009). Representatives from 12 viral families were detected, but unlike the two previous viromes that had been published at that time using comparable techniques, ssDNA viruses and large dsDNA viruses that putatively infect eucarya were the dominant viral types rather than bacteriophages. The ssDNA viruses were related to circoviruses, geminiviruses, nanoviruses and satellites; viruses previously only known to infect plants and animals indicating they are much more diverse than previously suspected and may constitute new viral families. Samples taken in summer showed a shift in the viral community composition towards phycodnaviruses similar to Ostreococcustauri virus, OtV5. This shift potentially reflects an increase in the host algae that are stimulated to bloom by the increased light availability.

Understanding of Antarctic viral diversity and ecology is still in its early days as a complete viral survey is problematic due to the lack of a universal viral gene or even universal genetic material. Furthermore, the enormous depth of viral diversity remains largely unsampled so most viral sequences have no significant similarity to sequence data repositories (López-Bueno *et al.*, 2009). What is clear is that viruses perform a crucial role in shaping community structure, driving host evolution, contributing to the dissolved nutrient pool, and understanding them is essential to understanding ecosystem function (**?**).

1.10 Objectives

This study aimed to use a primarily metagenomic approach complemented with microscopy and metaproteomics to gain an integrative understanding of whole ecosystems. Ace Lake and Organic Lake were chosen as the study

sites because as meromictic systems, differences in the microbial population can be examined across vertical gradients, they are also largely isolated, as marine-derived systems, adaptation of marine microbiota to lake system can be examined, they are sites of moderate diversity and may be reservoirs of unknown taxa. Using this methodology, not only can the taxonomic composition of the lakes be determined but also the functional potential of the microbial population and insight into the active members of the community. The objectives of the research were:

1. To develop complementary analyses to metagenomic analysis.

2. To determine the microbial and viral composition of the lake communities and their functional potential.

3. To identify the genes expressed by the community.

4. To reconstruct genomic information of dominant taxa of interest.

5. To integrate environmental and biological data to model the lake microbial interactions and geochemical processes.

Chapter 2

Metaproteogenomic analysis of Ace Lake

Co-authorship statement

Sections from this chapter 2 have been published as:

Federico M. Lauro, Matthew Z. DeMaere, **Sheree Yau**, Mark V. Brown, Charmaine Ng, David Wilkins, Mark J. Raftery, John A.E. Gibson, Cynthia Andrews-Pfannkoch, Matthew Lewis, Jeffery M. Hoffman,Torsten Thomas, and Ricardo Cavicchioli. An integrative study of a meromictic lake ecosystem in Antarctica. *International Society of Microbial Ecology Journal* 5:879–895, 2011.

I performed the metaproteomic mass spectra analysis, epifluorescence imaging, microbial and viral counts and wrote the corresponding sections of the publication. Only these parts of the publication are included in the results and discussion of this chapter.

Analyses performed by others that support the work presented in this chapter are as follows: Research was designed by Federico Lauro, Mark Brown, Torsten Thomas, John Gibson and Ricardo Cavicchioli. Sample collection was performed by Federico Lauro, Mark Brown, Torsten Thomas, Jeffery Hoffman and Ricardo Cavicchioli. DNA extraction and clone library preparation of 2006 samples was performed by Cynthia Andrews-Pfannkoch and Jeffery Hoffman of the J.Craig Venter Institute (JCVI). DNA sequenc-

ing quality control was performed by Matthew Lewis of the JCVI. Metagenomic sequence filtering, mosaic assembly and annotation was performed by Matthew DeMaere. Protein extraction, one-dimensional sodium dodecyl sulphate-polyacrylamide gel electrophoresis and liquid chromatography mass spectrometry performed by Charmaine Ng. Assistance in mass spectra analysis was provided by Mark Raftery.

Relation of this work to thesis objectives

2.1   Summary

2.2   Introduction

Ace Lake is a meromictic saline lake in the Vestfold Hills, Antarctica. It is the best studied of all the meromictic lakes in the Vestfold Hills and possibly Antarctica. Extensive physical, chemical and biological data has been collected from Ace Lake in the last decades.

Ace Lake is a highly stratified lake system, 25 m deep at its deepest point. It is ice-covered for approximately 9 months of the year and thaws some summers. Water is marine-derived and a largely neutral water balance has ensured salinity is close to that of seawater. Slightly fresher surface water of X can overlay the deeper waters when the ice-cover melts. The lake is physically separated into an aerobic mixolimnion, a steep chemocline/oxycline at 12.7 m and an anoxic monimolimnion below. The monimolimnion is sulfidic with accumulated methane. Initial diversity analysis of the sediment showed microbial diversity was reduced (Bowman *et al.*, 2000b).

As a physically and chemically well-characterised system of moderate diversity, Ace Lake was chosen as a model ecosystem to implement a high thtoughput metagenomic analysis. Samples were obtained down the depth profile at 5, 11.5, 12.7, 18 and 23 m depths corresponding to each of the three zones. Sampling was conducted as part of the global ocean sampling (GOS) expedition (Rusch *et al.*, 2007) by using size fractionation of microbial biomass onto 3.0, 0.8 and 0.1 μm membrane filters.

Microbial and viral community composition was assessed from the metagenomic dataset. Significant differences were found in taxonomic composition of each size fraction within each sample depth and stratification of the microbial community between the three zones of Ace Lake. The mixolimnion community is similar to a marine surface water assemblage consisting of a high abundance of the SAR11 clade of *Alphaproteobacteria* related to "*Candidatus* Pelagibacter ubique" and green algae of the order *Mamiellales*. However, diversity is reduced by one order of magnitude reduced (Lauro *et al.*, 2011). Unlike Southern Ocean surface water, the mixolimnion is overrepresented in *Cyanobacteria* related to *Synechococcus* and *Actinobacteria*, which may

represent taxa that mark the transition of a marine to lake community. A dense, near clonal population of green sulphur bacteria related to *Chlorobium* termed C-ace reside at the chemo/oxycline at 12.7 m (Ng *et al.*, 2010; Lauro *et al.*, 2011). Below, in the anoxic monimolimnion is a diverse, primarily heterotrophic community with abundant sulphate-reducing bacteria (SRB) and methanogenic *Archaea*.

Preliminary work on the metaproteome down the vertical profile of Ace Lake has been performed using the National Center for Biotechnology Information (NCBI) non-redundant database (NR) database (?). However, there were few protein identifications. Identification rate reduced as diversity of the sample increased. A focused metaproteogeonomic analysis conducted on the dense GSB layer using the matched metagenome as the database resulted in many more protein identifications compared to using the (Ng, 2010). Assessment of the genetic complement of Ace Lake showed a concurrent stratification of the functional potential in each zone of Ace Lake (Lauro *et al.*, 2011). Significantly, GSB appeared to be crucial in the lake ecosystem as they had the greatest genetic potential for nitrogen and carbon fixation as well as sulphur cycling(?Lauro *et al.*, 2011). Metaproteomic analysis was able to identify which proteins were actively expressed and thus the active pathways of the GSB metabolism which are so crucial to the function of the lake (Ng *et al.*, 2010).

This study aimed to expand on the metagenomic analysis of the water column of Ace Lake using complementary analyses. Metaproteomic analysis was used to identify expressed proteins of the 0.1 μm size fraction Ace Lake using a matched metagenomic databases for protein identification to infer which taxa and metabolic processes were active at time of sampling. To determine cellular and viral densities and validate the efficacy of size-fractionation with a modified epifluorescence microscopy procedure was developed and implemented.

2.3 Materials and methods

2.3.1 Ace Lake samples

Water samples were collected from Ace Lake (68°28′23.2″S, 78°11′20.8″E), Vestfold Hills, Antarctica on 21 and 22 December 2006. A 2 m hole positioned above the deepest point (25 m depth) of the lake was drilled through

the ice cover of Ace Lake to reach the lake surface. A volume of 1–10 L was collected by sequential size fractionation through a 20 μm pre-filter directly onto filters 3.0, 0.8 and 0.1 μm pore-sized, 293 mm polyethersulfone membrane filters (Rusch et al., 2007), along the depth profile as described previously (Ng *et al.*, 2010). Samples were taken in the order, 23, 18, 14, 12.7, 5 and finally, 11.5 m.

After samples from each depth were collected, the sample racks were sequentially washed with 2×25 L 0.1 N NaOH, 2×25 L 0.053% NaOCl, and 2×25 L fresh water. The sample hose was flushed with water from each depth before being applied to the filters. A *Chlorobium* signature was identified at 5 m, but not immediately above the GSB layer at 11.5 m. As the next sample taken after sampling at 12.7 m was at 5 m, and then 11.5 m, despite all equipment being thoroughly washed with bleach, NaOH and water, the simplest explanation for the GSB signature at 5 m is carry-over from sampling of the dense biomass at 12.7 m.

A sonde probe (YSI model 6600, YSI Inc., Yellow Springs, OH, USA) was used to record depth, dissolved oxygen content, pH, salinity, temperature and turbidity throughout the water column of the lake. Total organic carbon was determined using a total organic carbon analyzer, TOC-5000A (Shimadzu, Kyoto, Japan) equipped with a ASI-5000A auto sampler (Shimadzu), and particulate organic carbon by standard protocols (`http://www.epa.gov/glnpo/lmmb/methods/about.html`) at the Centre for Water and Waste Technology, UNSW.

2.3.2 DNA sequencing and data cleanup

DNA extraction and Sanger sequencing was performed on 3730xl capillary sequencers (Applied Biosystems, Carlsbad, CA, USA) and pyrosequencing on GS20 FLX Titanium (Roche, Branford, CT, USA) at the J.Craig Venter Institute in Rockville, MD, USA (Rusch *et al.*, 2007). The scaffolds and annotations will be available via Community Cyberinfrastructure for Advanced Microbial Ecology Research and Analysis (CAMERA) and public sequence repositories such as the NCBI and the reads will be available via the NCBI Trace Archive.

Sanger reads were trimmed according to quality clear ranges. The quality of pyrosequencing reads was assessed as follows: a basic local alignment search tool (BLAST) nucleotide database was created from the Sanger reads

of the 0.1 μm fraction of samples GS230, GS231 and GS232 (see Table of metagenomic data). After blasting the corresponding pyrosequencing reads against each database with a minimum bitscore of 80 and maximum e-value of 0.1, reads were binned according to length. The percentage of reads for each bin lacking a match to the Sanger read database was recorded. The percentage reads at least 25% repetitive after MDUST (**?**) analysis at default settings, and the percentage of reads containing N's, were assessed. In contrast to earlier pyrosequencers (**?**), no length-dependent bias in reads containing N's was observed. However, short reads had a disproportionately high number of repeats. Moreover, based on the proportion of reads with no match to the Sanger data set, both very short and very long reads had a disproportionately high number of errors; an observation that was previously reported (**?**).

On the basis of this analysis, a three step filtering process was applied to each sample. Reads were initially run through the Celera sffToCA (**?**) pre-processor followed by Lucy (**?**) and finally, excluding the bottom 8% and top 3% of reads determined from the read length distribution. As the sffToCA (v5.3) pre-processor removes all reads with a perfect prefix of any other read it overcomes the 'perfect duplicates problem (**?**).

After this process, <5% of the reads belonged to clusters of duplicates with three or more reads, and clusters of orthologous groups (COG) of proteins classification of these reads showed an over-representation of category L (replication, recombination and repair) that includes mobile genetic elements, which are often duplicated, suggesting a potential biological significance for the duplicated reads. It is possible these residual duplications are a result of high gene copy number or localized fragility of DNA sequences that might be biasing the shear points.

2.3.3 Metagenomic DNA assembly and annotation

Mosaic assemblies were generated for each sample fraction using Celera whole genome shotgun (WGS) Assembler v5.3 (**?**). For each assembly, the runtime parameters used were as outlined for 454 sequencing data in the published standard operating procedure
(http://sourceforge.net/apps/mediawiki/wgs-assembler/index.php?
title14SFF_SOP). As none of the samples can be considered clonal, these are regarded as stringent assemblies (Rusch *et al.*, 2007). Each 0.1 μm

fraction assembly was a hybrid of Sanger and 454 read data, wherein the estimated genome size was manually set to minimize the number of unitigs from abundant organisms being falsely classified as degenerate (Rusch *et al.*, 2007). Annotation of each sample fraction assembly was carried out using an in-house pipeline, wherein the pipeline stages consisted of genomic feature detection and subsequent annotation. Detected features consisted of open-reading frames (ORFs), transfer RNA and rRNA. Each detected ORF was further annotated by BLAST comparison against NR, Swissprot and Kyoto Encyclopedia of Genes and Genomes (KEGG)-peptide sequence databases and by biosequence analysis using profile hidden Markov models (HMMER) comparison against the Institute of Genomic Research curated protein database (TIGRFAM) (**?**), COG (**??**) and known marker genes (**?**). In all cases the cut-off e-value was a maximum of 1e−5.

2.3.4 Epifluorescence microscopy

Samples of unfiltered lake water and the flow-through from 3.0 and 0.8 μm filters from all depths were collected on November 2008 and fixed on site in formalin 1% (v/v). The samples were stored at -80°C for subsequent direct counts of cells and VLPs. Enumeration was performed according to the method of Patel *et al.* (2007) with modifications. Lake water samples were filtered onto 0.01 μm pore-size polycarbonate filters (25 mm Poretics, GE Osmonics, Minnetonka, MN,USA). Filters were air dried, then placed with the back of the filter on top of a 30 ml aliquot of 0.1% (w/v) molten low-gelling-point agarose and allowed to dry at 30°C. Samples were stained by the addition of 1 ml working solution (1 in 400 dilution in 0.02 μm filtered sterile Milli-Q) of SYBR Gold$^\circledR$ (Molecular Probes, Eugene, OR, USA) to 25 ml of mounting medium (VECTASHIELD HardSet, Vector Laboratories, Burlingame, CA, USA). Stained samples were counted immediately, or stored at -20°C for up to a week before counting. Samples were visualised under wide-blue filter set (excitation 460–495 nm, emission 510–550 nm) with an epifluorescence microscope (Olympus BX61, Hamburg, Germany).

2.3.5 Metaproteomic analysis

Proteins were extracted from membrane filters from all 0.1 μm fractions from the six depths (5, 11.5, 12.7, 14, 18 and 23 m), and one dimensional-sodium

dodecyl sulphate polyacrylamide gel electrophoresis (1D-SDS PAGE) and in gel trypsin digestion, liquid chromatography and mass mass spectrometry (MS), and two dimensional mass spectrometry (MS-MS) data analysis and validation of protein identifications performed as previously described (Ng *et al.*, 2010), with minor modifications. The specta generated were searched against the protein sequence database corresponding to that depth constructed from the 0.1 μm mosaic assemblies. The number of protein sequences in each database were as follows: 5 m, 138,208; 11.5 m, 133,948; 12.7 m, 27,142; 14 m, 62,436; 18 m, 71,512; and 23 m, 128,878. Scaffold (version Scaffold_2_05_01, Proteome Software Inc., Portland, OR, USA) was used to validate MS-MS-based peptide and protein identifications. Peptide and protein identifications were accepted if they could be established at >95% and 99% probability, respectively, as specified by the Peptide Prophet algorithm (**?**). Protein identifications required the identification of at least two peptides.

Proteins that contained similar peptides and could not be differentiated based on MS-MS analysis alone were grouped to satisfy the principles of parsimony and are referred to as a protein group. Spectral counting was used to semi-quantitively estimate protein abundance. The total assigned spectra that matched to each identified protein were exported from Scaffold 2.0. For similar proteins that have shared peptides (a protein ambiguity group), spectra were assigned to the protein with the most unique spectra. To normalize for variation in total spectra acquired between sample replicates, the number of spectra of each protein was multiplied by the average total spectra divided by the total spectra of the individual replicate. The spectral count of each protein was averaged across the replicates. As longer proteins are more likely to be detected, the average spectral counts were divided by the length of the protein. This value is equivalent to the normalized spectral abundance factor (**??**). In order to compare the relative abundance of proteins between depths, the normalized spectral abundance factor was divided by the average read depth of the contig (scaffold or degenerate) to which the protein mapped.

If >90% of a scaffolds length consisted of surrogate (highly degenerate unitig) sequence, the average read depth of the surrogate was used. For identified proteins that were part of a protein group the longest protein length and largest read depth value in the group was used. Pairwise comparisons

of each zone were conducted on COG assigned proteins. The normalized spectral counts from each protein was aggregated based on their COG annotation. All proteins that were part of an ambiguity group were confirmed to share the same COG annotation to ensure counts were not biased because of the common spectra.

The summed spectral counts from 5 and 11.5 m (mixolimnion), and 14, 18 and 23 m (monimolimnion) were pooled. Statistical significance of differences between each zone was assessed using Fisher's exact test, with confidence intervals at 99% significance calculated by the NewcombeWilson method and HolmBonferroni correction (P-value cutoff of 1e−5) in Statistical Analysis of Metagenomic Profiles (STAMP) (?). All proteins identified, including their gene identifier, normalized spectral abundance, COG and KEGG Orthology identifiers, KEGG locus tag and matching COG or KEGG description are provided in Supplementary Table S1.

2.4 Results and discussion

2.4.1 Development of epifluorescence microscopy methodology for cell and VLPs enumeration

As part of the integrative study of Ace Lake, an epifluorescence microscopy methodology was developed. Motivation for visualising microbiota from size fractionated water samples was twofold. Firstly, visualising microbiota from water samples allows examination of cellular morphologies and more importantly, enumeration of cells and VLPs. Cellular and VLP densities are not obtainable from the metagenomic data and necessitates a complementary method of determination. Determining viral densities is extremely important as the first studies to do so found that viruses are the most numerous biological entities on the planet and likely play a large role in plankton mortality in the ocean (Bergh *et al.*, 1989; Proctor and Fuhrman, 1990). VLP counts from marine environments vary with depth and trophic status ranging from 10^6 to 10^8 VLP ml^{-1} (?). Viral counts from various aquatic environments have differed from site to site indicating a variable role of virally mediated mortality.

Secondly, size fractionation of suspended microbial biomass from aquatic environments has been utilised as part of the landmark Sargasso Sea metagenomic study (Venter *et al.*, 2004) and subsequent GOS expedition (Rusch

et al., 2007). The Ace Lake samples were collected using the same sampling strategy as the GOS dataset but has sequence information from all three filter sizes. Thus, the visualisation of microbial morphologies from each stage of the filtration process would allow validation of the filtration process.

Developing a revised method was necessary due to inability to source 25 mm diameter 0.02 μm pore-size Anodisc filters (Whatman) that have been long used with fluorescent nucleic acid dyes for this purpose (**?**Noble and Fuhrman, 1998). They were marked for discontinuation in December 2008 after the take over of Whatman by GE Healthcare and was the cause of a global shortage that was strongly opposed by the viral ecology community (Torrice, 2009). Since conducting this research, production of 25 mm Anodisc filters has resumed, but the need for alternative methodolodies during the shortage was so great that alternative protocols were developed independently in other research groups (Budinoff *et al.*, 2011; Diemer *et al.*, 2012) stressing the importance in microbial ecology of epifluorescence microscopy for enumeration and the utility of alternative methodologies.

Clear polycarbonate Track Etch^{TM} (PCTE) filters were selected for use as a viable alternative product for the following reasons.

1. They have defined pore-size.

2. They are available in 0.015 μm pore-size allowing capture of VLPs.

3. They have previously been used for VLPs enumeration (**??**).

4. They have a long history of use with cellular enumeration (**?**) and can therefore be easily adopted for widespread use.

There are several disadvantages to using PCTE membranes over Anodiscs which are as follows.

1. Polycarbonate is not as robust as 25 mm diameter Anodisc filters that have a built in plastic support ring around the edge.

2. The polycarbonate of PCTE filters cannot be blotted dry as the Al_2O_e of Anodisc filters.

3. PCTE filters appear to have high background fluorescence.

4. PCTE filters have a much slower flow rate that Anodisc filters.

The protocol used is detailled in the materials and methods 2.3. The main disadvantages of using PCTE were overcome for the purposes required for this study. To counteract the fragility of the PCTE filters, vacuum filtration was not performed above X pressure limit and filters. No cracks or tears were observed during visualisation. PCTE filters also have a tendency to crinkle when mounted on the glass slide making visualisation of cells and VLPs on a single focal plane difficult. Agarose was used to embed the filters to help flatten the membrane and aid in mounting. However, this was not strictly necessary if filters were dried well and the membrane mounted carefully so that it was pressed flat against the glass slide.

The back ground fluorescence of the clear filters was low when stain is only incorporated into the mountant after filtration rather than staining in the column before filtration.

Filtration onto very small pore-size also necessitated a very strong seal of the filter column against the glass, but the slow flow-rate was not a property of the PCTE filters that could be overcome. As the filtration and visualisation was performed on fixed samples in the laboratory the time taken for filtration of 2–3 hours for each sample, this was not deemed problematic for this study.

Overall, a viable alternative methodology was successfully developed for visualisation and enumeration of cells and VLPs using PCTE membrane filters. To be a viable alternative to Anodisc-based methods in terms of accuracy of counts, a comparison of this methodolody and Anodisc-based protocols using viral samples of known densities needs to be performed. No Anodisc filters could be obtained for such a comparative analysis to validate its use in this study. However, for the purposes of this study, which is to show the relative differences in morphotypes between sample depths and size fractions, this method was more than suitable. Other groups have since shown it to be comparable.

2.4.2 Size and depth stratification of the community supported by cell and VLP densities

Development of fluorescence microscopy methodology using 0.01 μm pore-size polycarbonate filters for simultaneous cellular and viral counts shows:

1. Size fractionation procedure appeared effective.
2. Morphological differences supports stratification of the community.

3. Visualisation of the morphology supported the metagenomic data that saw size fractionated and taxonomically stratified community.

4. Virus to bacteria ratios can give important information about the community.

At 12.7m depth, the light levels, and the sharp transition in oxygen content and salinity (Fig. S2) favour the dominance of a very high-density (2.2×10^8 cells ml^{-1}) of a single type of GSB of the genus *Chlorobia*, refered to as C-Ace (Ng et al 2010). Viral signatures were essentially devoid in this zone. The ratio of bacteriophage to total viral population increased proportionally in the larger size fractions consistent with trophic analyses that indicate that the larger size fractions are mostly copiotrophic (Fig. S8) particle attached bacteria and therefore likely to be sensitive to lysogenic phage infection (Lauro et al 2009). The 23 m unfiltered lake water contained very high levels (1.3×10^8 VLPs ml^{-1}) of VLPs. The high diversity of bacteria and archaea in all size fractions of the monimolimnion (Fig. 2) is consistent with the presence of a high viral population (Rodriguez-Valera et al., NRM, 2009).

2.4.3 Development of metaproteomic methodology

1. Using a matched metagenome instead of NR for protein identification greatly increased the number of identifications. 1.1 Except at the bottom zone, likely because the community is too diverse so greater coverage of the metagenome is required. In parallel with taxonomic diversity increasing with depth (with the exception of the GSB layer), the rate of metaproteomic identification of proteins decreased with depth (Table S2). The majority of the proteins that were detected (e.g. 67% at 23 m) were for hypothetical proteins that tended to lack orthologs in well-characterized organisms, highlighting both the functional importance and novelty of this anaerobic zone of the lake.

2. More specific information could be assigned to the taxonomic groups such as. 2.1. The Actinobacteria sequences in the mixolimnion were associated with a diverse phylogenetic cluster (Luna cluster) mainly contributed by freshwater ultramicrobacteria (Hahn et al., 2003). Several Luna cluster isolates contain rhodopsin genes (Sharma et al., 2009) and similar gene sequences were present in the Ace Lake oxic zone data and found to be expressed (167820670 and 163154474; Table S2). 2.2.This is consistent with

the identification of clustered regularly interspaced short palindromic repeat (CRISPR) associated CRISPR-associated proteins (CAS) proteins Cse2, Cse3 and Cse4 (165526330, 165526332 and 165526334, respectively) in the 12.7 m metaproteome (Table S2). The CAS gene locus (cas3, cse1, cse2, cse3, cse4, cas5, cas1b), to which the proteins map, shares its organisation with CAS loci of sequenced GSB, and groups with the *E. coli* subtype/variant 2. The CRISPR/CAS system is likely to confer phage resistance to C-Ace, akin to the role in other organisms (Karginov and Hannon 2010; Horvath and Barrangou 2010).

3. Using Scaffold to validate protein identification and perform spectral counts was helpful. 3.1Same protein identifications as Charmaine except one or two. 3.2 Able to quantify differences between mixolimnion and monimolimnion. The diversity and abundance of ATP-binding cassette (ABC) transporters was lowest in the 0.1 μm fractions at 23 m (Fig. 3), and a correspondingly low number were detected in the metaproteome (Table S2). In contrast, numerous transporters, predominately ABC type, were identified in the metaproteome of the mixolimnion samples, with a high COG representation of transporters for carbohydrates ($\sim$34% of normalized spectra), amino acids ($\sim$32%) and inorganic ions ($\sim$9%) (Table S2 and Fig. S11). All transporters in the metaproteome were of bacterial origin and conservative phylum level assignments of the normalised spectra showed the majority to originate from *Proteobacteria* (69%), of which SAR11 comprised 46% and *Actinobacteria* 19% (Table S2). A high proportion of expressed genes with transport functions have also been reported for SAR11 from coastal (Poretsky et al. 2010) and open ocean waters (Sowell et al. 2009) (Morris et al. 2010?). Oligotrophs, such as SAR11 not only posses a low-diversity of high-affinity transporters (Lauro et al., 2009), but regulate the relative abundance of transporters expressed in response to dissolved organic carbon (DOC) availability (Poretsky et al. 2010). The prevalence of amino acid and simple sugar transporters (Table S2), and the low DOC concentration in the Ace Lake mixolimnion (Fig. 1) is likely to reflect efficient utilisation of these substrates from the DOC pool. Two SAR11 transport proteins that were detected in Ace Lake (Table S2) were not detected from the Sargasso Sea (Sowell,et al. 2009): an ectoine/hydroxyectoine (167807477 and 167892279) and a zinc ABC transporter (167933120). The zinc ABC transporter is likely to support zinc efflux in response to zinc concentrations

29

which are ∼70-fold higher in the mixolimnion of Ace Lake compared to seawater (Rankin 1999). Conversely, phosphate transporters were a major class detected from the Sargasso Sea (Sowell,et al. 2009) but were absent from the Ace Lake metaproteome; consistent with lower phosphate levels in the Sargasso Sea (<5 nM) compared to Ace Lake (1–12 μM). The differences in transporter expression between Ace Lake and oceanic SAR11 are likely to signify adaptive growth strategies that have evolved in the Ace Lake SAR11 community. The high numbers of bacteriophages in the monimolimnion (detected by microscopy, Fig. S5 and S6; metaproteomics, Table S2; metagenomics, Fig. 2), and increase in DOC observed at depth (Fig. 1), also indicates that carbon turnover in the monimolimnion is likely to be tightly coupled to the carbon flux going through a viral shunt, as proposed for open ocean systems (Suttle, C. A. Viruses in the sea. Nature 437, 356361 (2005)). The bacteriophages are also likely vehicles for mediating gene exchange. Most of the genetic potential to cycle the nitrogen pool appears to be limited to nitrogen assimilation throughout the lake and remineralisation in the monimolimnion (Fig. S14). The detection of glutamine synthetase (GlnA) and glutamate synthases (GltBS) in the metaproteome (Table S2) are supportive of active nitrogen assimilation. In the mixolimnion, GlnA was linked to SAR11 and Actinobacteria, and they are likely to be responsible for nitrogen absorption in the oxic zone. At the oxycline, GlnA and GltB from GSB were abundant (Table S2), indicating an important role for nitrogen assimilation at this zone in the lake. Genes for assimilatory sulphate reduction (ASR) were present in metagenome data of all three fractions at all depths, although they were lowest at the oxycline. However, there was no evidence for expression of the genes as ASR proteins were not detected by metaproteomics. In contrast, multiple subunits of the GSB dissimilatory sulfide reductase (DSR) complex were identified (Ng et al. 2010 and Table S2) indicating functionality of this pathway at the oxycline. GSB likely utilise the DSR system to convert sulphur to sulfite and the polysulfide-reductase-like complex 3 to oxidize sulfite to sulphate. SRB may then reform sulphide completing the sulphur cycle between the GSB and SRB (Ng et al. 2010). While SRB were detected at the three depths of the monimolimnion, sulphate is depleted in the water column and sediment at the bottom of the lake limiting their dissimilatory capacity (Rankin et al 1999). Finally, sulphate in the mixolimnion can be linked to sulphur-oxidation by SAR11 (Meyer and

Kuever, Microbiology 153:3478-3498) and a concomitant lack of capacity to perform sulphur reduction.

2.5   Conclusions

Using complementary approaches helps to validate the research methodology and metagenomic inferences about the whole community. Specifically, differences in size and depth was shown by both microscopy and metagenomics to be apparent. This both validates the method of size fractionation as a viable approach to broad separation of the community, as well as supports the assertion that there was a large difference in community at different depths. Using a matched metaproteomic database showed a huge increase in the number of protein identifications. This was provided that metagenomic coverage was good. Using a metaproteomics, genes identified as potentially relevant in the metagenome were found to be expressed, supporting their importance. For example, it showed the CRISPR genes were active and may be a defence against phage. It also showed Actinorhodopsins were expressed. It showed that abundant genes were normally abundant in the metaproteome, such as transport proteins that give insight into what substrates are important components of the DOC pool. New inferences could be drawn from the metaproteome, such as the preference for labile substrates such as active sulfate reduction which is not apparent from the sulphate concentration.

Chapter 3

Virophage control of Antarctic algal host–virus dynamics

Co-authorship statement

A version of this chapter has been published as:

Sheree Yau, Federico M. Lauro, Matthew Z. DeMaere, Mark V. Brown, Torsten Thomas, Mark J. Raftery, Cynthia Andrews-Pfannkoch, Matthew Lewis, Jeffery M. Hoffman, John A. Gibson, and Ricardo Cavicchioli. Virophage control of antarctic algal host–virus dynamics.
Proceedings of the National Academy of Sciences USA 108:6163–6168, 2011.

Contributions to this publication by other researchers is as follows. Research was designed by Federico Lauro, Mark Brown, Torsten Thomas, John Gibson and Ricardo Cavicchioli. Sample collection was performed by Federico Lauro, Mark Brown, Torsten Thomas, Jeffery Hoffman and Ricardo Cavicchioli. DNA extraction and clone library preparation of 2006 samples was performed by Cynthia Andrews-Pfannkoch and Jeffery Hoffman of the J. Craig Venter Institute. DNA sequencing quality control was performed by Matthew Lewis of the J. Craig Venter Institute. Metagenomic sequence filtering, global assembly and annotation was performed by Matthew De-Maere. Assistance in mass spectronomy and mass spectra analysis was pro-

vided by Mark Raftery. Assistance in analysis of Eucarya taxonomy was provided by Mark Brown. Analysis of virophage abundance over time was performed by Federico Lauro.

Apart from these contributions, I performed all other data analyses and interpretations.

3.1 Abstract

Viruses are abundant ubiquitous members of microbial communities, and in the marine environment affect population structure and nutrient cycling by infecting and lysing primary producers. Antarctic lakes are microbially dominated ecosystems supporting truncated food webs where viruses exert a major influence on the microbial loop. Here we report the discovery of a new virophage (relative of the recently described Sputnik virophage) that preys on phycodnaviruses that infect prasinophytes (phototrophic algae). By performing metaproteogenomic analysis on samples from Organic Lake, a hypersaline meromictic lake in Antarctica, complete virophage and near-complete phycodnavirus genomes were obtained. By introducing the virophage as an additional predator of a predator-prey dynamic model we determine that the virophage stimulates secondary production through the microbial loop by reducing overall mortality of the host and increasing the frequency of blooms during polar summer light periods. Virophages remained abundant in the lake two years later, and were represented by populations with a high level of major capsid protein sequence variation (25–100% identity). Virophage signatures were also found in neighbouring Ace Lake (in abundance), and in two tropical lakes (hypersaline and fresh), an estuary, and an ocean upwelling site. These findings indicate that virophages regulate host-virus interactions and influence overall carbon flux in Organic Lake, and play previously unrecognised roles in diverse aquatic ecosystems.

3.2 Introduction

It has been known for at least 20 years that viruses frequently infect and lyse marine primary producers causing up to 70% of cyanobacterial mortality (Proctor and Fuhrman, 1990; **?**). Eucaryotic phytoplankton are preyed upon by large dsDNA phycodnaviruses (PVs) causing bloom termination in globally distributed species (3,6). Elevated levels of dissolved organic carbon (DOC) (7) and numbers of heterotrophic bacteria (8-10) occur during algal blooms indicating that viral lysis of eucaryotic algae stimulates secondary production. Viruses also suppress host populations at concentrations below bloom-forming levels, with abundance being controlled by the efficiency and production rates of the infecting viruses (11, 12).

Antarctic lakes are microbially dominated ecosystems supporting few, if any metazoans in the water column (13, 14). In these truncated food webs, viruses are expected to play an increased role in the microbial loop (15). Low complexity Antarctic lake systems are amenable to whole community based molecular analyses where the role that viruses play in microbial dynamics can be unravelled (14). Attesting to this, a metagenomic study of Lake Limnopolar, West Antarctica uncovered a dominance of eucaryotic viruses and ssDNA viruses previously unknown in aquatic systems (16).

We established a metaproteogenomic program for Organic Lake (68°27′23.4″S, 78°11′22.6″E), which is located in the Vestfold Hills, East Antarctica, in order to functionally characterize its microbial community. Organic Lake is a shallow (7 m) hypersaline ($\sim$230 g L^{-1} maximum salinity) meromictic lake with a high concentration of dimethylsulphide ($\sim$120 μg $^{-1}$) in its anoxic monimolimnion (17, 18). Water temperature at the surface of the lake can vary from -14 to $+15$°C while remaining sub-zero at depth (19, 20). The lake is eutrophic, with organic material sourced both from autochthonous production and input from penguins and terrestrial algae. The high concentrations of organic material reflect slow breakdown in the highly saline lake water. The salt in the lake was trapped along with the marine biota when the lake was formed due to falling sea level c. 3,000 y BP (21, 22). The lake sediment has both low species diversity (Shannon-Weaver diversity: 1.01) and richness (Chao non-parametric index: 32 ± 12) (23). Unlike high latitude lakes, viral abundance has been reported to increase with trophic status (15) and with salinity in Antarctic lakes (24).

Here we report the analysis of the surface water of Organic Lake, highlighting the presence of a relative of the recently described Sputnik virophage, a small eucaryotic virus that requires a helper *Acanthamoeba polyphaga* mimivirus (APMV) to replicate (25). From metagenomic DNA, a complete Organic Lake virophage (OLV) genome was constructed (the second virophage genome to be described), and near-complete genomes of its probable helper Organic Lake phycodnaviruses (OLPVs).

3.3 Materials and methods

3.3.1 Samples and DNA sequencing

Water samples collected from Organic Lake were:

1. Surface water from the eastern side of the ice-free lake (68°27′25.48″S, 78°11′28.06″E) December 24, 2006.

2. A depth profile collected through a 30 cm hold drilled through the surface ice above the deepest point in the lake (68°27′22.15″S, 78°11′23.95″E), November 10, 2008.

3. surface water from the north-east side of the partially ice-covered lake (68°27′21.02″S 78°11′42.42″E), December 12, 2008.

Samples were sequentially filtered through a 20 μm pre-filter and biomass captured onto 3.0, 0.8 and 0.1 μm membrane filters as described previously (1, 2). The samples from 2008 also included 50% (v/v) RNAlater. DNA extraction, sequencing and quality validation was performed as previously described (1, 2). DNA sequencing was performed at the J. Craig Venter Institute in Rockville, MD, USA.

3.3.2 Transmission electron microscopy

Unfiltered Organic Lake surface water from December 24, 2006 (fixed on-site in 1% (v/v) formalin) was concentrated and a solvent exchange performed with sterile filtered ammonium acetate solution 1% (w/v) using a 50 kDa cut-off Microcon centrifugal filter device (Millipore) according to the manufacturers instructions. Formvar coated 200 mesh copper grids were floated on a droplet of sample for 30 min, excess liquid wicked off and the grid

negatively stained for 30 s with uranyl acetate 2% (w/v). The sample was visualised using a JEOL1400 transmission electron microscope at 100 kV at 150,000 to 250,000 × magnification.

3.3.3   Metagenomic assembly and annotation

Mosaic metagenomic assemblies were generated as previously described (1, 2). For the 0.1 μm Organic Lake 2006 sample, assembly was a hybrid of Sanger and 454 read data (Table S1). For all other sample size fractions, runtime parameters used were standard for 454 sequencing data. Low GC ($\geq$ 51%) scaffolds > 10 kb from the 0.1 μm 2006 assembly had high coverage (> 45 ×) indicating these were from the dominant taxa. One of these scaffolds was binned as virophage and the rest as PV.

To further separate the OLPV types and assess the completeness of their genomic content, highly conserved single copy PV orthologues were identified as follows. An all against all BLASTp search was conducted with protein sequences from the ten available PV genomes (*Acanthocystis turfacea Chlorella* virus 1, PbCV-1, PbCV AR158, PbCV FR483, PbCV NY2A, *Emiliania huxleyi* virus 86, *Ectocarpus siliculosus* virus 1, *Feldmannia* sp. virus, *Ostreococcus* virus 5, *Ostreococcus tauri* virus 1) and APMV (which was included as a close PV relative). BLASTp results were parsed and clustered using orthoMCL V1.4 (3, 4).

Pairs of each orthologue were located on eight of the PV scaffolds. The location of each orthologue pair had a complementary distribution so the eight scaffolds were able to be sorted unambiguously into two strains (OLPV-1 and OLPV-2). OLPV-1 ribonucleotide reductase α-subunit appeared as duplicated on different scaffold ends, likely as an artefact of its proximity to an assembly break point. The remaining high coverage scaffolds were searched for predicted proteins present in one OLPV strain but not in the other and assigned to the strain in which it was absent. Comparison of OLPV-1 and OLPV-2 scaffolds was performed using tBLASTn of concatenated scaffolds from each strain and visualised using the Artemis Comparison Tool (ACT) (5). DNA sequence data is available in Genbank and CAMERA (http://web.camera.calit2.net).

3.3.4 Organic Lake virophage genome completion and annotation

The high coverage (77 ×), large number of Sputnik homologues that encode essential functions and length of the putative OLV scaffold from the 0.1 μm 2006 hybrid assembly indicated it was a near-complete genome. Reads from this scaffold were reassembled at high stringency and visualised using Phred/Phrap/Consed (6) to complete the sequence. Mate-pair data indicated a circular molecule and primers were designed to span the ends of the scaffold and sequence across the gap (Table S5). Touch-down PCR was performed with eDNA from 0.1 μm 2006 sample, the product used for nested PCR and the final product was cloned and sequenced. The complete genome was manually annotated and visualised using Artemis (7). Translated ORFs (minimal size 120 amino acids) were compared (BLASTp) to GenBank, to the all metagenomic ORF peptide database on CAMERA (`http://web.camera.calit2.net`) and to predicted proteins from OLPV-1 and OLPV-2 scaffolds. Comparisons between the OLV genome and OLPV-1 /OLPV-2 were performed with tBLASTn and visualised using ACT (5).

3.3.5 Phylogenetic analysis

Translated amino acid sequences from viral marker genes of interest were retrieved from the 0.1 μm 2006 metagenomic assemblies from this study, GenBank and CAMERA all metagenomic reads ORF peptide database. Homologous sequences were aligned using MUSCLE v3.6 (8). Neighbour-joining analysis, test for clade support (bootstrap analysis 2000 replicates) and tree drawing was performed with Molecular Evolutionary Genetics Analysis (MEGA) software v4 (9). Maximum likelihood analysis (JTT substitution model) and test for clade support (aLRT analysis) was performed with PhyML (10) and the tree visualised using MEGA. 18S rRNA gene sequences were retrieved from reads of all filter sizes, compared (BLASTn, e-value < 1.0e−5) to the SILVA100 SSURef database, aligned and phylogeny performed using ARB as previously described (1, 2). The abundance and similarity of virophages in all lake samples and filter sizes was estimated using BLASTp (evalue < 1.0e−5) to search using the OLV MCP sequence against a database of proteins predicted from sequencing reads. The database was generated as previously described (1) and the percent

identity of the BLAST hit was used as a proxy for species similarity.

3.3.6 Metaproteomic analysis

Metaproteomics of proteins from the 0.1 μm filter from 2006 was performed as previously described (1, 2), with minor modifications. The protein sequence database was generated by combining ORFs from the 3.0, 0.8 and 0.1 μm mosaic assemblies with 130,581 sequences in the database. Scaffold 3.0 (Proteome Software Inc.) was used to validate MS/MS based peptide and protein identifications. Protein identification data is available in Table S2.

3.3.7 Model of algal host–virus and virophage dynamics

To model the effect a virophage would have on algal Pyramimonas algal host populations in Organic Lake, modified Lotka-Volterra equations were used describing the OLV as a predator of predator OLPV. The original equations are given by:

$$\frac{\mathrm{d}A}{\mathrm{d}t} = \alpha A - \varepsilon P A \tag{3.1}$$

$$\frac{\mathrm{d}P}{\mathrm{d}t} = \theta P A - \mu P \tag{3.2}$$

Where:

A is the number of *Pyramimonas* (prey).

P is the number of OLPV (predator).

α is the specific growth rate of the prey.

θ is the specific production rate of the predator.

ε is the rate of predator mediated death of prey.

μ is the specific decay rate of the predator.

Equation 3.1 describes the change in *Pyramimonas* abundance and equation 3.2 the change in OLPV abundance in the absence of OLV. In the presence of OLV, *Pyramimonas*, OLPV and OLV dynamics are described by the following equations:

$$\frac{\mathrm{d}P}{\mathrm{d}t} = \theta PA - \mu P - \omega PV \qquad (3.3)$$

$$\frac{\mathrm{d}V}{\mathrm{d}t} = \beta PV - \gamma V \qquad (3.4)$$

Where:

V is the number of teh OLV (predator of predator).

ω is the rate of OLV mediated reduction in OLPV infective particles.

β is the production rate of OLV.

γ is the decay rate of OLV.

Equation 3.3 is a modified version of equation 3.2 which includes the effect of OLV on the change in abundance of OLPV. Equation 3.4 describes the growth properties of OLV as a predator of OLPV. Values for the variables for the solution shown (Fig. 4) were as follows: initial prey (10), predator (1) and predator of predator (10) numbers, $\alpha = 0.1$, $\theta = 0.0015$, $\varepsilon = 0.01$, $\mu = 0.01$, $\omega = 0.01$, $\beta = 0.015$ and $\gamma = 0.15$. COmplex PAthway Simulator (COPASI) software (11) was used to simulate prey, predator and predator of predator dynamics using the deterministic (LSODA) method.

3.4 Results and discussion

3.4.1 Dominance of phycodnaviruses in Organic Lake

Water samples from Organic Lake were collected December 2006 and November and December 2008, and microbial biomass collected onto 3.0, 0.8 and 0.1 μm membrane filters as described previously (14). A large proportion of shotgun sequencing reads (96.2%) from the 0.1 μm size fraction of the 2006 Organic Lake metagenome (Table S1) had no significant hits to sequences in the RefSeq database (tBLASTx with e-value $< 1.0e-3$, minimum alignment length: 60 bp, minimum identity: 60%). The degree of assembly was high, with 77% of reads forming part of a scaffold, indicating the sample contained a few abundant taxa of minimal diversity. Forty-five scaffolds were longer than 10 kb; the five longest ranged from 70 to 171 kb. GC content and coverage were used to separate scaffolds into taxonomic groups (Fig. S1). A broad division was evident between low ($\leq$ 41%) and high ($\geq$

51%) GC scaffolds suggesting they constituted two taxonomic groups. All scaffolds in the high GC group that could be assigned contained phage homologues, as did the one exceptional low GC scaffold. The low coverage in the high GC group showed bacteriophages were not abundant in the 0.1 μm fraction. These scaffolds were not analyzed further. The low GC scaffolds with confident assignments contained sequences matching conserved PV or APMV proteins. These PV-related scaffolds comprised 60% of assembled reads demonstrating that OLPVs were numerically dominant in the 0.1 μm fraction. Transmission electron microscopy (TEM) revealed the presence of virus-like particles with the dimensions and structure typical of PVs (Fig. 1A).

Within the low GC group, scaffolds separated into a high coverage ($>$ 45$\times$) group, including the five longest scaffolds, and a low coverage ($<$ 22$\times$) group. Two of the scaffolds in the high coverage group and one in the low coverage group contained the PV marker DNA polymerase B (DPOB). The two high coverage DPOB share 76% amino acid identity and both share $\sim$57% identity to the low coverage DPOB. DPOB is single-copy throughout the nucleo-cytoplasmic large DNA virus (NCLDV) family to which PVs belong (26,27), demonstrating that the Organic Lake surface waters contained two closely related abundant PV types (DPOB1) and (DPOB2), and a more distantly related lower abundance type (DPOB3).

Phylogenetic analysis clustered Organic Lake DPOB with unclassified lytic marine PV isolates that infect the prymnesiophytes *Chrysochromulina ericina* (CeV1) and *Phaeocystis pouchetii* (PpV), the prasinophyte *Pyramimonas orientalis* (PoV) (4,28), and uncultured marine PVs related to APMV (29, 30) (Fig. 2A and Fig. S2). As the host range of PVs broadly correlates with DPOB phylogeny (31, 32), OLPV would infect prasinophytes or prymnesiophytes. The most probable host is the prasinophyte, *Pyramimonas* (no prymnesiophyte 18S rRNA gene sequences were present in any size fraction of the Organic Lake metagenome) (Fig. S3).

Supporting the presence of more than one PV, pairs of single-copy PV orthologues (ribonucleotide reductase alpha and beta subunits, VV A32R virion packaging helicase, PBCV1 A482R-like putative transcription factor, VV D5 ATPase and VLTF2 family transcription factor) were identified in the high coverage scaffolds that shared an average of 81% percent amino acid identity. Based on the positions of single copy genes on the scaffolds and the

percent identity between them, the high coverage scaffolds were grouped into two strains designated OLPV-1 and OLPV-2 according to their DPOB phylogeny (Fig. 2A and Fig. S2). The remaining high coverage scaffolds were assigned to either strain, resulting in two near-complete genomes of ∼300 kb each (Fig. 2C), that are within the range of other sequenced PV genomes (155-407 kb). In addition, several OLPV genomic fragments contained PV homologues in high coverage scaffolds that could not be confidently assigned to either strain.

Both OLPV strains contain a PpV-like major capsid protein (MCP) designated MCP1 and another unique MCP designated MCP2 (Fig. 2B and Fig. S4). Both OLPV MCP1s were identified in the metaproteome (Fig. 2C and Table S2) but MCP2 was not. In addition to MCPs, the metaproteome contained a range of abundant structural proteins and others more likely to be packaged in the virion (e.g. chaperone), that were expressed by OLPV-1, OLPV-2 and/or an OLPV genomic fragment (Fig. 2C and Table S2). These data suggest that MCP1 is the major structural protein, and that both OLPV-1 and OLPV-2 were in a productive cycle in the lake at the time of sampling.

3.4.2 Complete genome of an Organic Lake virophage

Sputnik is a small (50 nm) icosahedral satellite virus of mamavirus (a new strain of APMV). It was termed a "virophage" because co-infection with Sputnik is deleterious to the mamavirus, resulting in abnormal virions and a decrease in mamavirus infectivity (25). One 28 kb scaffold in the low GC high coverage group had six out of 38 predicted proteins homologous to Sputnik virophage proteins (Fig. 3 and Table S3), and one PV homologue. The scaffold had a low GC content (∼30%), similar to the Sputnik genome, and was larger in size (28 kb vs 18 kb for Sputnik). Using PCR and sequencing, the scaffold was found to represent a complete circular virophage genome (the Sputnik genome is also circular). Virus-like particles resembling Sputnik in size and morphology were identified by TEM (Fig. 1B).

Sputnik homologues present in the Organic Lake scaffold included the V20 MCP, V3 DNA packaging ATPase, V13 putative DNA polymerase/primase and others of unknown function (V9, V18, V21 and V32) (Fig. 3 and Table S3). The Organic Lake virophage (OLV) is distinct to Sputnik as proteins

43

share 27-42% amino acid identity (28% MCP identity). OLV proteins include OLV9, the homologue of Sputnik V20 MCP, and OLV8, a fusion of the uncharacterised V18 and minor virion protein V19 from Sputnik (Fig. 3 and Table S3). The large number of homologues, including genes that fulfill essential functions in Sputnik (V20, V3 and V13), indicate that OLV and Sputnik have physiological similarities.

3.4.3 Gene exchange between virophage and phycodnaviruses

As PVs are related to APMV (27) and are abundant in Organic Lake, it stands to reason that OLPV is the helper of OLV. In the OLV genome, OLV12 is a *Chlorella* virus-derived gene, indicating that gene exchange has occurred between OLV and PVs (the function of OLV12 is discussed below). Similar observations were made for Sputnik, which carries four genes (V6, V7, V12 and V13) in common with the mamavirus, indicative of gene exchange between the viruses and possible co-evolution (25). As the V6, V7, V12 and V13 proteins have been associated with virophage-helper specificity, we reasoned that the functional analogues in OLV would have highest identity to proteins from its helper virus, rather than Sputnik.

By comparing OLV and OLPV, a 7,408 bp region was identified in OLV encoding five proteins (OLV17-22) with identity (32–65%) to sequences in both OLPV-1 and OLPV-2 (Fig. 2C, Fig. S5 and Table S3). OLV20 and OLV13 are collagen triple-helix-repeat-containing proteins, analogous to Sputnik collagen-like proteins (V6 and V7) involved in protein-protein interactions in the APMV virus factory. Sputnik can replicate with either mamavirus or APMV as a helper, although coinfection rates are higher with the mamavirus. V6 is the only protein with higher identity (69%) to mamavirus than APMV (42%) (25). Since OLV20 has equivalent identity (63%) with OLPV-1 and OLPV-2, it appears that OLV may be capable of interacting with both OLPV strains. Also within the conserved region, OLV22, is a 141 aa protein of unknown function that only matches sequences from OLPV and the Global Ocean Sampling (GOS) expedition (Table S3). Similar to OLV22, Sputnik V12 is a small protein (152 aa) of unknown function with high identity to APMV, and both may mediate a specific helper-virophage interaction. Other genes in this region of OLV can be mapped to OLVP, including a putative transmembrane protein (OLV17) and paralogous phage tail fibre repeat containing proteins, OLV18 and OLV19. Analogous to the

collagen-like proteins, OLV19 and OLV20 probably facilitate interactions between helper and virophage.

OLV12 (which is unique to OLV) consists of a C-terminal domain present in conserved hypothetical *Chlorella* virus proteins and an N-terminal domain most closely related to class 3 lipases that may confer OLV selectivity to a PV. OLV12 may function similarly to the Sputnik V15 membrane protein in modifying the APMV membrane (25). The Sputnik V13 consists of a primase domain and SF3 helicase domain related to NCLDV homologues, involved in DNA replication. The helicase domain of OLV25 and V13 are similar, although the primase domain is more similar to a protein from *Ostreococcus lucimarinus*, implying a past association of OLV with a prasinophyte alga host.

Genes unique to OLV point to adaptations specific to its helper-host system. Most notably, OLV possesses a N6 adenine-specific DNA methyltransferase, as does OLPV. In OLPV-1, genes for a bacterial type I restriction modification (RM) system are adjacent to a gene encoding a type I methylase-S target recognition domain protein, and upstream of a DNA helicase distantly related to type III restriction endonuclease (RE) subunits. A large number of *Chlorella* virus genomes have both 5mC and 6mA methylation (33), and several contain functional RM systems (34). The prototype *Chlorella* virus PbCV-1 possesses REs packaged in the virion for degrading host DNA soon after infection (35). In contrast to OLV/OLPV, DNA methyltransferases are absent in both Sputnik and APMV, indicating that the N6 adenine-specific DNA methyltransferase has been selected in OLV to reduce endonucleolytic attack mediated by OLPV.

3.4.4 Role of virophage in algal host–phycodnavirus dynamics

The presence of the virophage adds an additional consideration to the microbial loop dynamics. In batch amoeba cultures, co-infection of amoeba with APMV and Sputnik causes a 70% decrease in infective AMPV particles and a 3-fold decrease in lysis (25). To test how OLV affects OLPV and host population dynamics, we modelled the OLV as an additional predator of a predator in a Lotka-Volterra simulation (Fig. 4). In the model, the effect of virophage is robust, with equilibrium solutions across a wide range of parameter values (Fig. 4 shows one equilibrium solution). By decreasing

the number of infective OLPVs, the presence of OLV shortens the recovery time of the host population (Fig. 4C) and shifts the orbit away from the axis (Fig. 4D). The model reveals that the virophage stimulates the flux of secondary production through the microbial loop by reducing overall mortality of the host algal cell following a bloom, and by increasing the frequency of blooms during the summer light periods. Antarctic lake systems have evolved mechanisms to cope with long light-dark cycles (14) and shortened trophic chains. In Organic Lake and similar systems, a decrease in PV virulence may be instrumental in maintaining stability of the microbial food web.

3.4.5 Ecological relevance of virophages in aquatic systems

Metagenomic analysis of Organic Lake samples taken two years later in November (when the lake was ice covered) and December 2008 (partially ice-free) revealed sequences with 99% amino acid identity to OLV MCP indicating persistence of OLV in the ecosystem (Fig. 5 and Table S4). In addition, sequences with lower identity (25–90%) were detected, particularly in December, demonstrating Organic Lake virophages are highly diverse but OLV remained the dominant type.

From surface water samples of nearby Ace Lake (meromictic, surface 2% salinity), a large number of sequences were obtained that matched both the OLV MCP (Fig. 5 and Table S4) and PVs (14). All Ace Lake size fractions contained matches to OLV MCP, some with high identity (80–100%) and the majority with greater variation (25–80% identity) (Fig. 5 and Table S4). In contrast to Organic Lake where the largest number of matches was to the 0.1 μm size fraction, the majority of Ace Lake sequences were from the larger fractions (Fig. 5 and Table S4). This indicates the Ace Lake virophages were associated with host cells during sampling, or possibly with helper viruses that are larger than the OLPVs.

Extending the OLV MCP search to the GOS data revealed matches (25-28% identity) to sequences from the hypersaline Punta Cormorant Lagoon (Floreana Island, Galapagos), an oceanic upwelling near Fernandina Island (Galapagos), Delaware Bay estuary (NJ, USA), and freshwater Lake Gatun (Panama) (Table S4). The phylogenetic analysis of a conserved 103 amino acid region of the MCPs revealed a number of clusters, with Sputnik clustering with virophage sequences from Ace Lake that had low identity (22%)

to OLV MCP (Fig. 5 and Fig. S4). To improve searches for virophages and better understand their physiology and evolution, it will be valuable to target more genomes (e.g. the Ace Lake 167858124 relative with 40% MCP identity to Sputnik) and determine which genes are core to virophages and what relationship exists between genome complement and MCP identity.

In view of the implications of the virophage modelling (Fig. 4), the abundance and persistence of OLV in Organic Lake (Fig. 5, Table S4), and the presence of diverse virophage signatures in a variety of lake systems (fresh to hypersaline), an estuary, an ocean upwelling site and a water cooling tower (Sputnik), our study indicates that numerous types of virophages exist and play a previously unrecognised role in regulating host–virus interactions and influencing ecosystem function in aquatic environments.

3.5   Acknowledgements

Chapter 4

Heterotrophic resourcefulness and unusual sulphur biogeochemistry in a hypersaline Antarctic Lake

Co-authorship Statement

A version of this chapter has been submitted as:

Sheree Yau, Timothy J. Williams, Federico M. Lauro, Matthew Z. DeMaere, Mark V. Brown, John Rich, John A.E. Gibson, Ricardo Cavicchioli. Heterotrophic resourcefulness and unusual sulfur biogeochemistry in a hypersaline lake. _ISME Journal_ submitted, 2013.

Contributions to this manuscript by other researchers is as follows.

Research was designed by Federico Lauro, Mark Brown, John Gibson and Ricardo Cavicchioli. Sample collection was performed by Federico Lauro, Mark Brown and Ricardo Cavicchioli. Metagenomic sequence filtering, global assembly and annotation was performed by Matthew DeMaere. Assistance in intrepretation of functional potential provided by Timothy Williams.

Apart from these contributions, I performed all other data analyses and interpretations.

4.1 Abstract

Organic Lake is a shallow, marine-derived hypersaline lake in the Vestfold Hills, Antarctica that has the highest reported concentration of dimethylsulphide dimethylsulphide (DMS) in a natural body of water. To determine the composition and functional potential of the microbial community and learn about the unusual sulphur chemistry in Organic Lake, shotgun metagenomics was performed on size fractionated samples collected along a depth profile. Eucaryal phytoflagellates were the main photosynthetic organisms. Bacteria were dominated by the globally distributed heterotrophic taxa *Marinobacter*, *Roseovarius* and *Psychroflexus*. Candidate division RF3 was overrepresented at the oxycline and OD1 in the lake bottom. The dominance of heterotrophic degradation coupled with low fixation potential indicates possible net carbon loss. However, abundant marker genes for aerobic anoxygenic phototrophy, sulphur oxidation, rhodopsins and CO oxidation were also linked to the dominant heterotrophic bacteria and indicate use of photo- and lithoheterotrophy as mechanisms for conserving organic carbon. Similarly, a high genetic potential for the recycling of nitrogen compounds likely functions to retain fixed nitrogen in the lake. dimethylsulphoproprionate (DMSP) lyase genes (*dddD*, *dddL* and *dddP*) were abundant indicating DMSP is a significant carbon and energy source. Unlike marine environments, DMSP demethylases (*dmdA*) were less abundant than DMSP lyases indicating that DMSP cleavage is the likely source of the high DMS concentration. DMSP cleavage, photoheterotrophy, lithoheterotrophy and nitrogen remineralisation by dominant Organic Lake bacteria are potentially important adaptations to nutrient constraints. In particular, photo- and lithoheterotrophy reduces the extent of carbon oxidation for energy production allowing more carbon to be used for biosynthetic processes. The study sheds light on how the microbial community in Organic Lake has adapted to the unique physical and chemical properties of this Antarctic Lake environment.

4.2 Introduction

Due to the polar light cycle, phototrophic growth in Antarctic environments is relatively high in summer and negligible in winter (Laybourn-Parry *et al.*, 2005) and requires microbial life to survive under long periods under a scarcity of resources. To overcome this limitation, Eucaryotic phytoflagellates in Ace Lake engage in carbon mixotrophy by grazing on bacterioplankton to supplement their carbon requirements in the winter (Laybourn-Parry *et al.*, 2005). Marine heterotrophic bacteria are known to be similarly resourceful by exploiting light energy through photoheterotrophy that includes aerobic anoxygenic photosynthesis (AAnP) or via use of rhodopsins, or lithoheterotrophy such as oxidation of carbon monoxide (Moran and Miller, 2007). Heterotrophic bacteria that can harness energy sources apart from organic carbon can direct a greater proportion of carbon towards growth, which serves to conserve fixed carbon within a closed systems (Moran and Miller, 2007).

Organic Lake is shallow (6.8 m) and has variable surface water temperatures (-14 to $+15°$C) while remaining sub-zero throughout most of its depth (Franzmann *et al.*, 1987b; Gibson *et al.*, 1991; Roberts *et al.*, 1993; Gibson, 1999). The salt and marine biota in the lake originate from seawater that was trapped in a basin $\sim$3,000 BP (Zwartz *et al.*, 1998; Bird *et al.*, 1991). The bottom waters of Organic Lake are unusual due to the absence of hydrogen sulphide and the high concentration of the volatile gas dimethylsulphide (DMS) (Deprez *et al.*, 1986; Franzmann *et al.*, 1987b; Gibson *et al.*, 1991; Roberts and Burton, 1993; Roberts *et al.*, 1993). Concentrations of DMS as high as 5,000 nM have been recorded in Organic Lake (Gibson *et al.*, 1991), 100 times the maximum concentration recorded from seawater in the adjacent Prydz Bay and at least 1 000 times that of the open Southern Ocean (**?**). More than forty years ago atmospheric DMS was proposed to have a regulatory effect on global cloud cover as it is a precursor of cloud condensation nuclei (Lovelock and Maggs, 1972; Charlson *et al.*, 1987). However, the first enzymes involved in DMS production were only identified in the last six years (Todd *et al.*, 2007). Rapid progress has been made in this short period and the pathways and organisms involved in DMS transformations have been extensively reviewed (Johnston *et al.*, 2008; Schäfer *et al.*, 2010; Curson *et al.*, 2011; Reisch *et al.*, 2011; Moran *et al.*, 2012).

The main source of DMS in the marine environment is from the break-down of dimethylsulphoproprionate (DMSP). Eucaryal phytoplankton, in particular diatoms, dinoflagellates and haptophytes, produce large quantities of DMSP, an organosulphur compound that is thought to function principally as an osmolyte. DMSP is released due to cell lysis, grazing or leakage and follows two known fates: DMSP cleavage by DMSP lyases (DddD, -L, -P, -Q, -W and -Y) or demethylation by DMSP demethylase (DmdA). Both pathways are associated with diverse microorganisms that can utilize DMSP as a sole carbon and energy source. However, it is only the cleavage pathway that releases volatile DMS that can lead to sulphur loss through ventilation to the atmosphere. The very high levels of DMS in Organic Lake make it an ideal system for identifying the microorganisms and the processes involved in DMS accumulation.

The previous Organic Lake metagenomic study examined viruses from the 0.1 μm fraction of surface water that was collected from Organic Lake in December 2006, and November and December 2008 (Yau *et al.*, 2011). In the present study we focused on the cellular population rather than viruses. Our study determined the composition and functional potential of Organic Lake microbiota and, in conjunction with historic and contemporary physico-chemical data, generated an integrative understanding of the whole lake ecosystem.

4.3 Materials and methods

4.3.1 Characteristics of the lake and sample collection

The water level of Organic Lake was measured by surveying as +1.886 m relative to the survey mark (NMV / S / 53) located at 68°27′23.4″S, 78°11′22.6″E. Water was collected from Organic Lake on 10 November 2008 through a 30 cm hole in the 0.8 m thick ice cover above the deepest point in the lake. The sampling hole was established at 68°27′22.2″S, 78°11′23.9″E) following bathymetry measurements constructed on a metric grid. Samples were collected for metagenomics, microscopy and chemical analyses at 1.7, 4.2, 5.7, 6.5 and 6.7 m depths (maximum lake depth 6.8 m).

For metagenomics, lake water was passed through a 20 μm pore size pre-filter, and microbial biomass captured by sequential filtration onto 3.0 μm, 0.8 μm and 0.1 μm pore size 293 mm polyethersulfone membrane fil-

ters, and samples immediately preserved in buffer and cryogenically frozen in liquid nitrogen, as described previously (Ng *et al.*, 2010; Lauro *et al.*, 2011). Between 1–2 L of lake water was sufficient to saturate the holding capacity of the filters. deoxyribonucleic acid (DNA) was extracted from the filters, samples sequenced using the Roche GS-FLX titanium sequencer, and reads processed to remove low quality bases, assembled and annotated, as previously described (Ng *et al.*, 2010; Lauro *et al.*, 2011). A summary of the 2.4 Gbp of metagenomic data isprovided in Supplementary Table S1.

4.3.2 Physical and chemical analyses

An *in situ* profile of pH, conductivity, turbidity, dissolved oxygen (DO) and pressure was measured using a submersible probe (YSI sonde model V6600). A temperature profile was measured using a maximum-minimum mercury thermometer as the YSI probe did not have a capacity to record temperature below -10^{circ}C. The 5.7 m sample corresponded to the turbidity maximum and the 6.5 m sample to the turbidity minimum. Conductivity at *in situ* temperature was converted to conductivity at 15°C as described previously (Gibson, 1999). The adjusted conductivity brings the temperature to within a range suitable for estimating practical salinity using the formula of Fofonoff and Millard (1983). Salinity was likely to have been underestimated as it is higher than the range (2–42) for which the conductivity–salinity relation holds. However, the relative difference in salinity between the samples would be accurate.

Density was calculated from the *in situ* conductivity and temperature using the equations described by Gibson *et al.* (1990) and expressed at temperature T as:

$$\sigma_T = (1000 - density) kgm^{-3} \qquad (4.1)$$

Ammonia, nitrate, nitrite, total nitrogen (TN), total dissolved nitrogen (TDN), dissolved reactive phosphorus (DRP), total phosphorus (TP), total dissolved phosphorus (TDP), total organic carbon (TOC), dissolved organic carbon (DOC), total sulphur (TS) and total dissolved sulphur (TDS) were determined by American Public Health Associations Standard Methods at the Analytical Services, Tasmania. Values for dissolved nutrients were measured after filtration through a 0.1 μm pore size membrane filter. All

other nutrients were measured from water collected after filtration through the on-site 20 μm pore size pre-filter.

Ammonia, nitrate, nitrite, DRP, TN, TDN, TP and TDP were measured in a Flow Injection Analyser (Lachat Instruments, Colorado, USA). TOC and DOC were determined in the San++ Segmented Flow Analyser (Skalar, Breda, Netherlands). TS and TDS were analyzed in the 730ES Inductively Coupled Plasma–Atomic Emission Spectrometer (Agilent Technologies, California, USA). Principal Component Analysis (PCA) was performed using the PRIMER Version 6 statistical package (Clarke and RN, 2006) on the normalized physical and chemical parameters.

4.3.3 Epifluorescence microscopy

Water samples collected for microscopy were preserved in formaldehyde (1% v/v). Cells and VLP were vacuum filtered onto 25 mm polycarbonate 0.015 μm pore-size membrane filters (Nuclepore Track-etched, Whatman, GE Healthcare, USA) with a 0.45 μm pore-size backing filter. The 0.015 μm filter was mounted onto a glass slide with ProLong Gold$^{\textregistered}$ anti fade reagent (Invitrogen, Life Technologies, NY, USA) and 2 μl (25 dilution in sterile filtered milliQ water < 0.015 μm) SYBR$^{\textregistered}$ Gold nucleic acid stain (Invitrogen, Life Technologies, NY, USA). Prepared slides were visualized in an epifluorescence microscope (Olympus BX61, Hamburg, Germany) under excitation with blue light (460–495 nm, emission 510–550 nm). Cell and virus-like particle (VLP) counts were performed on the same filter over 30 random fields of view.

4.3.4 Cellular diversity analyses

Diversity of *Bacteria*, *Archaea* and *Eucarya* was assessed using small subunit ribosomal RNA (SSU) gene sequences. Metagenomic reads that matched the 16S and 18S ribosomal RNA (rRNA) genes were retrieved using Metaxa (Bengtsson *et al.*, 2011). Only sequences longer than 200 bp were accepted for downstream analysis.

The Quanititative Insights Into Microbial Ecology (QIIME) pipeline (version 1.4.0) (Caporaso *et al.*, 2010) implementing UCLUST, was used to group SSU sequences into operational taxonomic units (OTUs) at 97% percent identity against the SILVA SSU reference database (release 108) (www.

`arb-silva.de`). SSU sequences that did not cluster with sequences from SILVA were allowed to form new OTUs (no suppression). A representative sequence from each OTU was chosen and classified to the genus level using QIIME implementing the Ribosomal Database Project (RDP) classifier (Wang *et al.*, 2007) trained against SILVA. Assignments were accepted to the lowest taxonomic rank with bootstrap value $leq 85\%$.

To allow comparison of the relative abundance of taxa, the number of SSU matches per sample filter was normalized to the average number of reads (403 577). Statistical analysis on the relative SSU abundances was performed using the PRIMER Version 6 package (Clarke and RN, 2006). The SSU counts of each sample filter were aggregated to the genus level and square root transformed to reduce the contribution of highly abundant taxa. A resemblance matrix was computed using Bray-Curtis similarity. The upper mixed zone (1.7, 4.2 and 5.7 m) and deep zone (6.5 and 6.7 m) samples were designated as separate groups and an analysis of similarity analysis of similarity (ANOSIM) performed to test for difference between the two groups. BEST analysis was performed with the abiotic variables: conductivity, temperature, turbidity, DO, pH, TOC, TN, TP, TS, total C:N, total C:P, total N:P, cell counts and VLP counts. The Bio-Env procedure in BEST looks at all the abiotic variables in combination and finds a subset sufficient to best explain the biotic structure. A heat map with bi-clustering dendogram was generated using R and the package seriation (Hahsler *et al.*, 2007) on the normalized square-root transformed SSU counts.

4.3.5   Analysis of functional potential

The relative abundance and taxonomic origin of functional marker genes was used to determine the potential for carbon, nitrogen and sulphur conversions. The open-reading frames (ORFs) were predicted from trimmed metagenomic reads using MetaGene (Noguchi *et al.*, 2006) accepting those >90 bp in length. Open reading frames ORFs were translated using the standard bacterial/plastid translation table and compared to protein sequences from the Kyoto Encyclopedia of Genes and Genomes (KEGG) Genes database (release 58) using the basic local alignment search tool (BLAST) (Altschul *et al.*, 1990).

The BLAST output was processed using KEGG Orthology Based Annotation System (KOBAS) version 2.0 (**?**) accepting assignments to KEGG

Orthology (KO) groups with e-value $<1e-05$ and rank >5. KO groups used as functional markersare listed in Supplementary Table S2. Marker enzymes were assigned to taxonomic groups based on the species of origin of the best KEGG Genes BLASTp match.

Marker genes not represented by a KO group were assessed by BLASTp queries of marker gene sequences with experimentally confirmed function (Supplementary Table S3) against a database of translated ORFs predicted from metagenomic reads. Matches were accepted if the e-value was $<1e-10$ and sequence identity was within the range shared by homologs of the query sequence(s) (Supplementary Table S3). Matches to marker genes were normalized to 100 Mbp per sample and counted. Normalized frequencies of markers from the same pathway were averaged and those from different pathways were summed.

The same marker genes and BLAST procedure was used to compare the DMSP catabolism and photoheterotrophy potential of Organic Lake with nearby Ace Lake (Lauro *et al.*, 2011), Southern Ocean (SO) (Wilkins *et al.*, 2012) and global ocean sampling (GOS) metagenomes (Rusch *et al.*, 2007). Counts of single copy gene *recA* were also determined to estimate the percentage of genomes containing each marker gene (percentage of marker genes relative to *recA*). Matches to *recA* were accepted with e-value $<1e-20$ according to the cut-off established by Howard *et al.* (2008). For GOS samples, the BLAST database was generated from peptide sequences retrieved from Community Cyberinfrastructure for Advanced Microbial Ecology Research and Analysis (CAMERA) (camera.calit2.net) while the other BLAST databases were produced as for Organic Lake. The total number of trimmed base pairs for GOS samples was estimated by multiplying the number of reads from each sample by the average read length (822 bp) (Rusch *et al.*, 2007).

Marker gene sequences for phylogenetic analysis were clustered using the CD-HIT web server (Huang *et al.*, 2010) at 90% global amino acid identity. A representative sequence from the clusters that resided within a desired conserved region and homologs from cultured strainswere used in phylogenetic analyses performed in MEGA 5.05 (Tamura *et al.*, 2011) implementing MUSCLE with default parameters (gap opening penalty: -2.9, gap extension penalty: 0). Neighbor-joining was used to compute the phylogenies with a Poisson substitution model, uniform rates of change and complete deletion of alignment gaps. Node support was tested with bootstrap analysis (500

replicates).

4.4 Results and discussion

4.4.1 Abiotic properties and water column structure

In situ physico-chemical profiles (Supplementary Figure S2) measured over the deepest point in the lake (Supplementary Figure S3) determined the existence of two zones: an upper mixed zone above 5.7 m and a suboxic deep zone below 5.7 m (Figure 1A). The separation of the two zones was indicated by a pycnocline and oxycline starting at 5.7 m. The pH also decreased with DO, likely due to fermentation products such as acetic, formic and lactic acids that have been reported in the bottom waters (Franzmann *et al.*, 1987b; Gibson *et al.*, 1994). The deep zone was not completely anoxic (Supplementary Figure S2). Oxygen may be episodically introduced to bottom waters as a result of currents of cold dense water sinking during surface ice-formation (**?**). In comparison to meromictic lakes such as Ace Lake that have strong pycnoclines and a steep salt gradient in the anoxic zone, Organic Lake is shallow and has relatively weak stratification (Gibson, 1999).

Samples were collected from the upper mixed (1.7, 4.2 and 5.7 m) and deep (6.5 m and 6.7 m) zones. All nutrients, except for nitrate and nitrite reached maximum concentrations at 6.5 m (Table1) suggestive of a layer of high biological activity above the lake bottom. Consistent with this, cell and VLP counts were highest at 6.5 m. However, turbidity was lowest at this depth demonstrating turbidity was not principally determined by cell density (Figure 1B). Microscopy images did not show a shift in cell morphology that could account for the large drop in turbidity (Supplementary Figure S4), which suggests particulate matter primarily contributed to turbidity readings. The low turbidity and peak in cell counts and nutrients at the oxycline at 6.5 m may be caused by an active microbial community degrading particulate matter. This inference is supported by the report of high concentrations of dissolved organic acids and free amino acids in the deep zone (Gibson *et al.*, 1994) as these nutrients are indicative of the breakdown of high molecular weight carbohydrates, lipids and proteins. Furthermore, the C:N and C:P ratios throughout the lake were high compared to the Redfield ratio (**?**) except at 6.5 m indicating this was the only depth where dissolved nitrogen and phosphorus were not relatively limited (Table 1).

PCA analysis of physico-chemical parameters showed all samples, except the 6.5 m sample, separated with depth along the PC1 axis (Supplementary Figure S5). Accordingly, turbidity, TS and cell density were the strongest explanatory variables for the separation of the 6.5 m sample from the other deep sample, indicating that increased activity at 6.5 m was related to breakdown of particulate matter and sulphur chemistry.

4.4.2 Overall microbial diversity

SSU genes (3,959 reads) that were retrieved from the metagenome data grouped into 983 OTUs. OTUs for *Bacteria* comprised 76.2%, *Eucarya* 16.3% and 7.5% of SSU sequences could not be classified. Only 2 reads, assigned to a deep sea hydrothermal clade of *Halobacteriales* (Supplementary Table S4), were assigned to *Archaea* indicating they were rare in Organic Lake.

The most abundant bacterial classes, *Gammaproteobacteria*, *Alphaproteobacteria* and *Flavobacteria*, were represented by OTUs on all filter sizes at all depths (Figure 2A) and each consisted of one dominant genus, *Marinobacter*, *Roseovarius* and *Psychroflexus*, respectively (Figure 2C). Essentially all OTUs for *Cyanobacteria*/chloroplasts were classified as chloroplasts (Figure 2A), except for three reads that could not be assigned to any lower rank (Supplementary Table S4) indicating free-living *Cyanobacteria* were rare or absent. OTUs for moderately abundant bacterial classes were *Actinobacteria*, *Deltaproteobacteria*, *Epsilonproteobacteria*, and candidate divisions OD1 and RF3. Lower abundance divisions included OTUs for *Bacilli*, *Clostridia*, *Spirochaetes*, *Lentisphaeria*, TM7, *Opitutae*, *Verrucomicrobia*, Bhi80-139, Bd1-5, SR1 and *Chlamydiae* (Figure 2A).

The dominant eucaryal OTUs were for photosynthetic *Chlorophyta* (green algae) and *Dictyochophyceae* (silicoflagellate algae) (Figure 2B) principally assigned to the genus *Dunaliella* and the order *Pedinellales*, respectively (Supplementary Table S4). Lower abundance eucaryal OTUs included *Bacillariophyta* (diatoms), *Dinophyceae*, *Fungi* and heterotrophic *Choanoflagellida* and *Ciliophora* (see Supplementary Table S4 for lower taxonomic rank assignments).

4.4.3 Variation of microbial composition according to size and depth

Community composition varied with size fraction and depth. This was supported by seriation analysis that showed samples clustered according to size fraction, and those clusters further separated into upper mixed and deep zone groups (Figure 3). A significant difference in genus-level composition between the upper mixed and deep zones was supported by ANOSIM test (Rho: 0.53, significance: 0.1%). Differential vertical distribution of taxa is consistent with partitioning of ecological functions in the lake and in association with the physical and chemical data, described functional roles of those taxa.

20–3.0 μm fraction community composition

The upper mixed zone samples had a relatively high OTU abundance of *Dunaliella* chloroplasts and chlorophyte algae consistent with large active photosynthetic organisms concentrating near surface light. They are likely the main source of primary production in Organic Lake and have previously been reported to be the dominant algae (Franzmann *et al.*, 1987b). The SSU sequences for these algae at the bottom of the lake are likely to be due to sedimentation of dead cells or resting cysts.

Psychroflexus OTUs were overrepresented in the surface and 6.7 m samples. Consistent with enrichment on the 3.0 μm filters, *Psychroflexus* (formerly *Flavobacterium*) *gondwanensis* (Bowman *et al.*, 1998) isolated from Organic Lake (Franzmann *et al.*, 1987b) had cells 1.5–11.5 μm in length (Dobson *et al.*, 1991). *Flavobacteria* associate with phytoplankton blooms in the Southern Ocean (Abell and Bowman, 2005a,b; **?**), and have specialized abilities to degrade polymeric substances from algal exudates and detritus (reviewed in **?**, (**?**)). It is likely that Organic Lake *Psychroflexus* fills a similar ecological role. In support of this, *Psychroflexus* OTUs cluster with *Dunaliella* chloroplasts in the seriation analysis (Figure 3) and *P. gondwanensis* abundance in Organic Lake has been correlated with average hours of sunshine per day indicating population dynamics that is related to summer algal blooms (**?**). The *Psychroflexus* OTUs in the deep zone are most likely due to sedimentation as *P.gondwanensisis* non-motile and strictly aerobic (Dobson *et al.*, 1991).

Roseovarius OTUs were enriched at 4.2 m and 6.5 m suggesting different ecotypes may be present in the upper mixed zone compared to the deep zone. *Roseovarius tolerans*, an isolate from Ekho Lake in the Vestfold Hills, Antarctica has a cell size (1.1–2.2 μm; (Labrenz *et al.*, 1999)) that would be expected to be captured on the 0.8 μm filter. The *Roseovarius* captured on the 3 μm filter may therefore be a different species, or a strain similar to *R. tolerans* from Ekho Lake that exhibits different growth characteristics (i.e. larger cell size or forms aggregates). A strain of this species from Ekho Lake is capable of microaerophilic growth (Labrenz *et al.*, 1999). Overrepresentation at 6.5 m may therefore be indicative of growth at that depth rather than sedimentation because sinking cells would be more abundant close to the lake bottom at 6.7 m. *Roseovarius* OTUs cluster with *Dunaliella* chloroplast and *Psychroflexus* OTUs in the seriation analysis (Figure 3), suggesting that Organic Lake *Roseovarius* may be utilising compounds released from algal-derived particulate matter, or made available by processing of complex organic matter by *Psychroflexus*. *Roseovarius* is a member of the *Roseobacter* clade, which is inferred to have an opportunistic ecology frequently associated with nutrient-replete plankton aggregates, including by-products of flavobacterial exoenzymatic attack (**??**). Additionally,the diverse metabolic capabilities of the *Roseobacter* clade include DMSP degradation, AAnP and CO oxidation (reviewed in **?**). All of these capabilities should facilitate growth in both the upper mixed and deep zones of Organic Lake (see 4.4.5).

3–0.8 μm size fraction community composition

On the 0.8 μm filter, OTUs for *Marinobacter* dominated at all depths except 6.5 m. Their capture on this size fraction is consistent with the cell size of isolates (1.2–3 μm) (**?**). The genus is metabolically versatile, which likely permits it to occupy the entire water column. *Marinobacter* is heterotrophic and the genus includes hydrocarbon-degrading strains (e.g., **??**, although deep-sea metal-oxidising autotrophs have also been reported (**?**). Some isolates are capable of interacting with diatoms (**?**) and dinoflagellates (**?**). *Marinobacter* isolates from Antarctic lakes are capable of anaerobic respiration using dimethylsulphoxide (DMSO) (Matsuzaki *et al.*, 2006) or nitrate (Ward and Priscu, 1997). Analysis of functional potential linked to *Marinobacter* revealed additional metabolic capabilities potentially related to its dominance in Organic Lake (see Carbon resourcefulness in dominant

heterotrophic bacteria and Molecular basis for unusual sulphur chemistry below).

OTUs for RF3 and Halomonas were overrepresented at 6.5 m, and RF3 sequences were more abundant (Figure 2 and 3). Their relative abundance in the deep zone indicates a role in microaerophilic processes. The majority of RF3 sequences to date are from anaerobic environments including mammalian gut (**???**), sediment (**??**), municipal waste leachate (**?**), anaerobic sludge (**????**), a subsurface oil well head (**?**), and the anaerobic zone of saline lakes (**??**Bowman *et al.*, 2000a). However, some members have been found in surface waters (**???**) suggesting not all members are strict anaerobes.

Several *Halomonas* isolates have been sourced from Organic Lake including two described species *Halomonas subglaciescola* and *H. meridiana*, both of which grow as rods with dimensions consistent with capture on this size fraction (Franzmann *et al.*, 1987a; **?**). Despite these isolates being aerobic, *Halomonas* has been reported to be enriched at the oxycline in Organic Lake (**?**) indicating *Halomonas* in the lake plays an ecological role in the suboxic zone. This capacity may be linked to the ability of free amino acids and organic acids, which are abundant in the deep zone (Gibson *et al.*, 1994), to stimulate the growth of isolates (Franzmann *et al.*, 1987a).

0.8–0.1 μm size fraction community composition

A large number of eucaryal sequences were evident in the 0.1 μm size fraction. The upper zone was overrepresented by OTUs for *Pedinellales* (silicoflagellate algae) that co-varied with chloroplasts (Figure 2 and 3). *Pedinellales* have only been detected in Antarctic lakes from molecular studies (Unrein *et al.*, 2005; Lauro *et al.*, 2011) including Organic Lake (Yau *et al.*, 2011), and light microscopy studies of Antarctic Peninsula freshwater lakes reported 5–8 μm diameter cells resembling *Pseudopedinella* (Unrein *et al.*, 2005). It is possible that in Organic Lake small (0.80.1 μm) free-living members or chloroplast-containing cyst forms (**?**) exist. However, without evidence to support this (e.g. by microscopy) it seems more likely that the lake sustains a relatively small number of active photosynthetic cells and the sequences detected arise from cysts or degraded cellular material.

OTUs for *Candidatus* Aquiluna, in the Luna-1 cluster of *Actinobacteria* (**??**) were most abundant at 1.7 m. The genus has small cells (<1.2 μm; (**?**),

accounting for their concentration on this size fraction. Although originally described in freshwater lakes, the same clade was detected in abundance in Ace Lake (Lauro *et al.*, 2011) and surface Artic seawater (**?**) demonstrating that they play ecological roles in polar saline systems. In Ace Lake surface waters they were associated with utilisation of labile carbon and nitrogen substrates (Lauro *et al.*, 2011), and in Organic Lake surface waters they probably perform similar functions. The presence of this clade in the deep zone implies a facultative anaerobic lifestyle or sedimented cells.

The bottom of the water column was distinguished by the presence of OTUs for candidate divisions OD1 and TM7. OD1 was more abundant, and its prevalence on this size fraction is consistent with similar findings for size fractionation of ground water (**?**). OD1 is consistently associated with reduced, sulphur-rich, anoxic environments (**??**). OD1 from Zodletone Spring, Oklahoma, was reported to possess enzymes related to those from anaerobic microorganisms (**?**). Genomic analyses identified OTUs for OD1 in the anoxic zone of Ace Lake (Lauro *et al.*, 2011). The distribution of OD1 in Organic Lake is consistent with an anaerobic metabolism and potential involvement in sulphur chemistry.

4.4.4   Organic Lake functional potential

To determine the potential for functional processes in Organic Lake, gene markers for carbon, nitrogen and sulphur conversions (Figure 4) were retrieved from metagenomic reads. BEST analysis showed that variation in the population structure was significantly correlated (Rho: 0.519, significance: 0.3%) with the abiotic parameters, DO, temperature, TS and TN. The DO gradient has an obvious effect of separating aerobic from anaerobic taxa, and allows oxygen sensitive nitrogen and sulphur processes to occur in the deep zone. Functional potential, taxonomic composition and the physico-chemical data were integrated to infer the carbon, nitrogen and sulphur cycles.

4.4.5   Carbon resourcefulness in dominant heterotrophic bacteria

In both the upper mixed and deep zones, potential for carbon fixation was much lower than for degradative processes, indicating potential for

net carbon loss (Figure 4A). Potential for carbon fixation via the oxygen-tolerant Calvin cycle (Figure 4A) was originally assessed by presence of the marker genes ribulose-bisphosphate carboxylase oxygenase (RuBisCO) and phosphoribulokinase (*prkB*) (**?**). The majority of RuBisCO homologs were related to *Viridiplantae* (Table 2) supporting the ecological role of green algae as the principle photosynthetic organisms. RuBisCO was only associated with a small proportion of *Gammaproteobacteria* (Table 2), principally from sulphur-oxidising *Thiomicrospira*, indicating some *Gammaproteobacteria* are autotrophs. However, the majority of *prkB* matched to *Gammaproteobacteria* (Table 2), predominantly *Marinobacter*. Although deep-sea, iron-oxidising autotrophic members of *Marinobacter* have been isolated (**?**), all genomes reported for *Marinobacter* have *prkB* but lack RuBisCO genes. Across *Marinobacter* genomes the *prkB* homolog is consistently adjacent to a gene for a putative phosphodiesterase, suggesting that the enzymes expressed by these genes may be involved in a pathway involved in pentose phosphate metabolism unrelated tocarbon fixation. Albeit exceptional, this decoupling of *prkB* from RuBisCO involved in carbon fixation (forms I and II), also observed in *Ammonifex* (**?**), undermines the utility of *prkB* as a marker gene for the Calvin cycle within certain groups. Thus, there is no evidence for autotrophy in Organic Lake mediated by *Marinobacter*.

Evidence for carbon fixation via the reverse tricarboxylic acid (rTCA) cycle was also indicated (Figure 4A), with genes for ATP citrate lyase (*aclAB*) linked to sulphur-oxidising *Epsilonproteobacteria* (Table 2). In general, the rTCA cycle is restricted to anaerobic and microaerophilic bacteria (**?**), which is consistent with the detection of *Epsilonproteobacteria* in the lake bottom where oxygen is lowest, and the microaerophilic/anaerobic metabolisms characteristic of the group (**?**). Anaerobic carbon fixation was represented by potential for the Wood-Ljungdahl; or reductive acetyl-CoA (WL) pathway (Figure 4A). WL-mediated carbon fixation, for which CO dehydrogenase/acetyl-CoA synthaseis the key enzyme, was linked to*Firmicutes* and *Deltaproteobacteria* that are known to grow autotrophically using this pathway (**?**).

Potential for carbon loss by via respiration was indicated by an abundance of cytochrome C oxidase genes (*coxAC*) throughout the water column. In the deep zone, potential for fermentation was greatest at 6.5 m (Figure 4A) and likely the main biological activity that was occurring at that depth.

Fermentation was indicated by the marker gene lactate dehydrogenase (*ldh*). These genes were linked to *Firmicutes* (Table 2), which was only present at 6.5 m and represented by the classes *Clostridia* and *Bacilli* (Figure 2A). As the related candidate division RF3 (Tajima1999) also has relatively high abundance in this zone (Figure 2A) (see 0.8–3.0 μm size fraction community composition above), there is circumstantial evidence that RF3 possesses fermentative metabolism and may therefore play an important ecological role in Organic Lake by degrading high molecular weight compounds to organic acids that other organisms can utilize. Assimilation of fermentation products appears to play a greater role in Organic Lake rather than complete anaerobic oxidation involving methanogens or sulphate-reducing bacteria; the former were absent and the latter were present in low abundance (Figure 2A, 2C).

Alphaproteobacteria, predominantly *Roseovarius* (Figure 2C), were implicated in CO oxidation (Table 2; Supplementary Figure S6A), which is used to generate energy for lithoheterotrophic growth (**?**), although CO oxidation may also be involved in anaplerotic C fixation (**?**). The CO oxidation capacity was at a maximum at 6.5 m (Figure 4A), and therefore associated with the deep-zone *Roseovarius* ecotype of Organic Lake. CO oxidation can function as a strategy to limit oxidation of organic carbon for energy so that a greater proportion can be directed towards biosynthesis (**?**).

Photosynthesis reaction center genes *pufLM*, involved in photoheterotrophy via AAnP, were abundant in Organic Lake (Figure 4A; Table 2). These were linked to the *Roseobacter* clade of *Alphaproteobacteria* (Table 2), major contributors to AAnP in ocean surface waters (**??**). This is consistent with the known metabolic potential of bacteriochlorophyll A (BchlA) producing *Roseovarius tolerans* from Ekho Lake (Labrenz *et al.*, 1999). Photoheterotrophy can also be rhodopsin-dependent, with proteorhodopsin (PR) of marine *Flavobacteria* and *Vibrio* previously linked to light-dependent energy generation to supplement heterotrophic growth, particularly during carbon limitation (Gómez-Consarnau *et al.*, 2007, 2010). However, the function(s) of rhodopsins are diverse, and PRs are also hypothesized to be involved in light or depth sensing (Fuhrman *et al.*, 2008).

Rhodopsin genes were abundant in Organic Lake (Figure 4A), and were associated with all the dominant Organic Lake aerobic heterotrophic lineages (Supplementary Figure S6). Phylogenetic analysis revealed six well-

64

supported Organic Lake rhodopsin groups (Supplementary Figure S6). All groups had an L or M residue at position 105 (vs the SAR86 PR), denoting tuning to surface green light (Man *et al.*, 2003; Gómez-Consarnau *et al.*, 2007), and is characteristic of oceanic coastal samples (Rusch *et al.*, 2007). Four of the groups clustered with homologs of genera detected in the lake, namely *Marinobacter*, *Psychroflexus*, *Octadecabacter* and "*Ca.* Aquiluna" (Supplementary Figure S6 and Table S4). Another group (SAL-R group) originates from the sphingobacterium *Salinibacter ruber*, which produces xanthorhodopsin (Balashov *et al.*, 2005); it is therefore likely that Organic Lake *Sphingobacteria* (Supplementary Table S4) were the origin of this rhodopsin group. The most abundant group (OL-R1; Supplementary Figure S6) had no close homologs from GENBANK, but it was abundant on the 3.0 μm fraction and has a distribution suggesting it originates from Organic Lake members of the *Roseobacter* clade (Figure 4A). All ORFs adjacent to OL-R1 rhodopsin containing scaffolds wererelated to *Octadecabacter* further supporting their *Roseobacter* clade provenance (Supplementary Figure S7). Genes downstream of OL-R1 were involved in carotenoid synthesis, indicating OL-R1 is a xanthorhodopsin, occuring as a retinal protein or in a carotenoid complex (Balashov *et al.*, 2005).

Photoheterotrophic potential of Organic Lake was compared with other aquatic environments including nearby Ace Lake, SO and GOS expedition samples. The Organic Lake 0.1 μm fraction had the lowest rhodopsin counts and percentage of rhodopsin containing cells of all size-matched samples surveyed (Table 3). Non-marine GOS samples from the 0.1 μm fraction have been noted to have lower rhodopsin abundance (Sharma *et al.*, 2008), which was similarly evident from our analysis (Table 3). In contrast, the 3.0 μm Organic Lake size fractions had higher rhodopsin counts than Ace Lake and comparable counts to the SO samples, although the percentage of rhodopsin containing cells was still lower than that of the SO. The paucity of rhodopsins in the Organic Lake 0.1 μm fraction is likely due to the lack of SAR11 clade, which is expected to be the main source of rhodopsin genes in Ace Lake and marine samples. This indicates that although Organic Lake has an overall lower frequency of rhodopsin genes compared to sites for which size fraction-matched metagenomes are available, the rhodopsins associated with larger or particle-associated cells are as abundant as in the marine environment.

Counts of *pufLM* genes inthe Organic Lake 0.1 μm size fraction were similar to GOS sample, except for Punta Cormorant hypersaline lagoon which had the highest *pufLM* counts and percentage of AAnP cells (Table 3). However, the highest overallcounts of *pufLM* were from the 3.0 μm size fraction of Organic Lake, likely due to the high proportion of members of the *Roseobacter* clade. Notably, *pufLM* genes were not detected in high abundance in Ace Lake or the SO samples, indicating AAnP is a unique adaptation in Organic Lake among these polar environments. The similarly high abundance of *pufLM* genes in Punta Cormorant hypersaline lagoon indicates AAnP may be advantageous in environments with salinity above marine levels.

The contribution of light-driven energy generation processes to the carbon budget is difficult to infer from genetic potential alone. For example, the relative abundance of AAnP and PR genes in Arctic bacteria has been reported to be the same in winter and summer (Cottrell and Kirchman, 2009). Furthermore, regulation of pigment synthesis is complex; for example, BchlA expression in *R. tolerans* occursin the dark but is inhibited by continuous dim light (Labrenz *et al.*, 1999). However, it is possible that the apparent negative balance in carbon conversion potential could be ameriolated by photoheterotrophy performed by bacterial groups that are abundant in Organic Lake. In particular, the Organic Lake *Psychroflexus* could play a particular role as it has a PR related to *Dokdonia*, which was shown to function under carbon-limitation (Gómez-Consarnau *et al.*, 2007). Furthermore, detection of higher AAnP potential in Organic Lake than other aquatic environments linked with taxa known to be capable of AAnP, suggests it may have a greater influence in the carbon budget of Organic Lake.

4.4.6 Regenerated nitrogen is predominant in the nitrogen cycle

Nitrogen cycling potential throughout the lake was dominated by assimilation and mineralisation/assimilation pathways (Figure 4B). Glutamate dehydrogenase (GDH) genes (*gdhA*) were abundant (Figure 4B), and linked predominantly to *Alpha-* and *Gammaproteobacteria* and to a lesser extent *Bacteroidetes* (Table 2). However, the functional significance of the readily reversible GDH depends on its origin; *Bacteroidetes* are likely to use GDH in the oxidative direction for glutamate catabolism (**?**), whereas the use of GDH in the oxidative or reductive directions by *Proteobacteria* is likely

to depend upon the source of reduced nitrogen (ammonia vs amino acids). Glutamine synthetase (*glnB*) and glutamate synthase genes (*gltBS*), were predominantly linked to *Alpha-* and *Gammaproteobacteria* (Table 2), indicating the potential for high-affinity ammonia assimilation by these groups in Organic Lake. The high ammonia concentration in the deep zone (Figure 1B; Table 1) would result from a higher rate of mineralisation (ammonification) than assimilation. This is consistent with abundant OTUs for *Psychroflexus*(*Bacteroidetes*) in this zone, and due to either turnover of organic matter or lysis of *Bacteroidetes* cells after sedimentation in anoxic water. In addition, the gene for ammonia-generating nitrite reductase (*nrfA*) was linked to *Bacteroidetes* and *Planctomycetes* (Table 2), indicating ammonia may also be produced by these putative aerobic heterotrophs. Overall, the data suggest that ammonia is actively assimilated in the aerobic upper mixed zone, but is permitted to accumulate in the anaerobic deep zone.

Potential for nitrogen conversions typically found in other aquatic environments was greatly reduced in Organic Lake. There was a very low potential for nitrogen fixation that was confined to the deep zone (Figure 2B) and principally linked to anaerobic *Epsilonproteobacteria* (Table 2). This diazotrophic potential may not be realized by *Epsilonproteobacteria*, given the high ammonia concentration present in the deep zone. No ammonia monooxygenase genes (*amoA*) were detected.The potential for ammonia oxidation was only represented by hydroxylamine/hydrazine oxidase-like (*hao*) genes, which werein low abundance and linked to *Deltaproteobacteria* (Table 2). *hao* genes are present in non-ammonia-oxidising bacteria (**?**), and those from Organic Lake belong to a family of multiheme cytochrome c genes present in sulphate-reducing *Deltaproteobacteria* that have no proven role in ammonia oxidation. In the genomes of sulphate-reducing *Deltaproteobacteria* the*hao* gene is invariably situated adjacent to a gene for a NapC/NirT protein, which suggests a role in dissimilatory nitrate reduction. Collectively these data indicate an inability for nitrification to occur in the upper mixed zone and no potential for ammonia loss in the deep zone.

Denitrification genes (*norCB* and *nozB*) and genes for nitrate assimilation (*nasA*) were present throughout the water column (Figure 4B) and were linked primarily to *Gammaproteobacteria* (Table 2). Low nitrate and nitrite in the deep zone (Figure 1B, Table 1) indicates oxidized nitrogen has been depleted by dissimilatory or assimilitory reduction by heterotrophic

Gammaproteobacteria. Denitrification genes are phylogenetically widespread and usually induced by low oxygen or oxidized nitrogen species (**?**) and thus expected to be active in the deep zone or oxycline. However, denitrification may be inhibited even if conditions appear appropriate. For example, in Lake Bonney, Antarctica, denitrification occurs in the west lobe, but not in the east lobe of the lake despite the presence of anoxia, nitrate and denitrifying *Marinobacter* species (Ward and Priscu, 1997; Ward *et al.*, 2005). Moreover, in the absence of nitrification, denitrification and nitrate assimilation would be limited by the lack of potential to re-form oxidized nitrogen. The preponderance of assimilation/mineralisation pathways geared towards reduced nitrogen appears to reflect a "short circuit" of the typical nitrogen cycle that would conserve nitrogen in a largely closed system. Hence, the predominant nitrogen source is regenerated fixed nitrogen. Similar findings were also made for Ace Lake, although in this system the presence of a dense layer of green sulphur bacteria with the potential to fix nitrogen augments the nitrogen cycle (Lauro *et al.*, 2011).

4.4.7 Molecular basis for unusual sulphur chemistry

Several meromictic hypersaline lakes in the Vestfold Hills,including Organic Lake, with practical salinity >150 are characterized by an absence of hydrogen sulphide and photoautotrophic sulphur bacteria (Burke and Burton, 1988). Although sulphate is present (Franzmann *et al.*, 1987b), geochemical conditions of these lakes are not conducive to dissimilatory sulphur cycling between sulphur oxidising and sulphate reducing bacteria typical of other stratified systems such as Ace Lake (Ng *et al.*, 2010; Lauro *et al.*, 2011). Consistent with this, potential for dissimilatory sulphate reduction represented by dissimilatory sulfite reductase (*dsrAB*) and adenylylsulphate reductase (*aprAB*)linked to sulphate-reducing *Deltaproteobacteria* (Table 2) was lowin Organic Lake. Sulphate-reduction potential was confined to the 6.7 m sample (Figure 4C) where oxygen concentration was lowest and *Deltaproteobacteria* were present (Figure 2A).

Capacity for oxidation of reduced sulphur compounds, represented by the sulphur oxidation multienzyme genes (*soxAB*), was present throughout the water column (Figure 4C) and linked primarily to *Alpha-* and*Gammaproteobacteria* (Table 2). Sulphur-oxidising *Alpha-* and*Gammaproteobacteria* are known to oxidize sulphur compounds, such as thiosulphate, aerobically. Although

a small proportion of *Gammaproteobacteria* had the capacity for autotrophy (see 4.4.5), the majority of sulphur-oxidizers were likely chemolithoheterotrophs as they were related to heterotrophic *Marinobacter* and *Roseobacter* clade. The sulphur dehydrogenase genes *soxCD* linked to *Alpha-* and *Gammaproteobacteria* were similarly present throughout the water column. *soxCD* are accessory components of the Sox enzyme system without which complete oxidation of thiosulphate cannot occur (Friedrich *et al.*, 2005). Thus the presence of *soxCD* indicates complete oxidisation likely occurs, although the different distribution of *soxAB* and *soxCD* in the water column (Figure 4C) suggests a proportion of the community may lack *soxCD* and deposit sulphur.

Sulphur-oxidising *Epsilonproteobacteria* possessing *soxAB* genes (Table 2) were present only in the deep zone of Organic Lake (Figure 2A,C) and were related to autotrophic deep sea sulphur-oxidizers, some members of whichare capable of anaerobic sulphur oxidation using nitrate (Yamamoto and Takai, 2011). It is unlikely that appreciable sulphur oxidation occurs in the deep zone as the known terminal electron acceptors, oxygen and nitrate, are deplete and the abundance of sulphur oxidising *Epsilonproteobacteria* is low (Figure 2A). *Epsilonproteobacteria* were also linked to a capacity for oxidation of sulphide to elemental sulphur by utilising sulphide:quinone oxidoreductase (*sqrA*) (Figure 4A, Table 2). In this pathway, sulphur is released as polysulphides, which is a potential biological source of the abundant polysulphides that have been detected in the lake *Roberts1993b*.

It is likely that the limited anaerobic dissimilatory sulphur cycle contributes to the accumulation of DMS in Organic Lake in the deep zone. In the upper mixed zone, DMS could potentially be oxidized as a carbon and energy source or utilized as an electron donor by sulphur-oxidising bacteria (Schäfer *et al.*, 2010). In anoxic zones, methanogenic *Archaea* or sulphate-reducing bacteria are the main organisms known to breakdown DMS (Schäfer *et al.*, 2010). Methanogens and genes involved in methanogenesis were not detected, nor has methane been detected (Gibson *et al.*, 1994) leaving sulphate-reduction the most likely route of DMS catabolism. The low dissimilatory sulphate reduction potential in the deep zone coupled with the relatively stagnant waters would likely minimize DMS oxidation and loss by ventilation. DMS would therefore be expected to accumulate in the deep zone if production rates were higher than breakdown.

To determine the source of high DMS in the bottom waters of Organic Lake, the genes involved in DMS formation were surveyed. Genes for DMSP lyases *dddD*, *dddL* and *dddP*, were detected in Organic Lake at levels comparable to other dominant processes such as respiration and fermentation (Figure 4C) indicating DMSP is an important carbon and energy source in Organic Lake. *dddD* was the most abundant of the Organic Lake DMSP lyases (Table 3) and comprised two main types: MAR-dddD and OL-dddD (Supplementary Figure S8). Neither of these types clustered with the non-functional *Dinoroseobacter shibae* DFL 12 and *Ruegeria pomeroyi* DSS-3 *dddD* homologs (Todd *et al.*, 2011) or carnitine coenzyme A transferase outgroups, thereby providing support for their proposed role as functional DMSP lyases. The MAR-dddD type includes the *Marinobacter* sp. ELB17 *dddD* homolog, and MAR-dddD sequences were most abundant on the 0.8 μm fraction where *Marinobacter* OTUs were also abundant, indicating MAR-dddD derives from Organic Lake *Marinobacter* (Supplementary Figure S8). OL-dddD did not have a close relative from cultured bacteria making its precise taxonomic origins uncertain. The abundance of OL-dddD on the 3.0 μm fraction suggests it originates from *Alphaproteobacteria*. OL-dddD containing contigs carried genes of mixed *Alpha-* and *Gammaproteobacterial* origin supporting its provenance from one of these classes and consistent with the "pick n' mix" arrangement of genes found beside sequenced *dddD* regions (Johnston *et al.*, 2008). Adjacent to OL-dddD was *dddT* (Supplementary Figure S9), a betaine, choline, carnitine transporter (BCCT) family protein that likely functions in substrate import, demonstrating OL-dddD forms an operon-like structure,similar to *Halomonas* sp. HTNK1 (Todd *et al.*, 2010).

Two *dddL* groups were detected in Organic Lake: SUL-dddL and MAR-dddL (Supplementary Figure S10). The former includes the *Sulfitobacter* sp. EE-36 *dddL* and the latter the *Marinobacter manganoxydans* MnI7-9 homolog indicating they originate from *Roseobacter* clade and *Gammaproteobacteria*, respectively. *Sulfitobacter* sp. EE-36 has demonstrated DMSP lyase activity and the *dddL* gene alone is sufficient for DMS generation (Curson *et al.*, 2008). These data indicate that the Organic Lake members of the SUL-dddL group perform the same functional role. The MAR-dddL clade appears to be an uncharacterized branch of the *dddL* family. emphdddP was detected as the least abundant of the DMSP lyases (Table 3). Phylogenetic analyses showed Organic Lake *dddP* likely originate from *Roseovarius* (Sup-

plementary Figure S11). The Organic Lake sequences formed a clade with the functionally verified *Roseovarius nibinhibens* ISM *dddP* (Todd *et al.*, 2009).

A single type of DMSP demethylase, *dmdA* was identified. It clustered with *Roseobacter* clade *dmdA* (Supplementary Figure S12), corresponding to the marine clade A (Howard *et al.*, 2006), and includes the functionally verified (**?**) DSS-3 homolog. These data indicate that the Organic Lake sequences correspond to true DMSP demethylases and not related glycine cleavage T proteins or aminomethyltransferases (Howard *et al.*, 2006). DMSP cleavage appears to be a significant source of DMS in Organic Lake. DMSP likely originates from *Bacillariophyta* or *Dinoflagellida*as Organic Lake *Dunaliella* have been reported not to produce DMSP in culture (Franzmann *et al.*, 1987b). Based on the abundance of marker genes, DMSP cleavage is predicted to occur at highest levels in the deep zone (Figure 4C) where the DMS concentration has been measured to be highest (Deprez *et al.*, 1986; Franzmann *et al.*, 1987b; Gibson *et al.*, 1991; Roberts and Burton, 1993; Roberts *et al.*, 1993). DMS can also be produced in anoxic environments from the reduction of DMSO, degradation of sulphur containing amino acids, and sulphide methylation (Schäfer *et al.*, 2010). Our data indicate that some DMSO reduction linked to *Firmicutes* could occur, but is not likely a major pathway (Figure 4C), and the potential for the other DMS yielding processes could not be determined because the enzymes involved in these pathways have not been established. When cultivated, Halomonas isolates from Organic Lake produced DMS from cysteine (Franzmann *et al.*, 1987b) providing some evidence that DMS production from anaerobic degradation of amino acids can occur. Abiotic pathways for anaerobic production of DMS have also been proposed (Roberts *et al.*, 1993).

The potential for DMSP cleavage was more than twice that of DMSP demethylation (Figure 4C). This is unusual compared to the marine environment or Ace Lake where DMSP demethylation potential is much higher than cleavage (Table 3). Previous estimates have similarly shown marine environments to have demethylation potential up to two orders of magnitude higher than cleavage (Howard *et al.*, 2008; Todd *et al.*, 2009; **?**; Reisch *et al.*, 2011). The frequency of DMSP lyase genes *dddD* and *dddL* in Organic Lake exceeded those of all other environments, except Punta Cormorant hypersaline lagoon, where dddL abundance was comparable (Table 3). This

suggests selection in Organic Lake for DMSP cleavage due to functional advantage and/or selection for taxa that carry DMSP lyase genes. There is evidence that high DMSP cleavage potential is adaptive in hypersaline systems, as a high proportion of ddd genes were similarly detected in Punta Cormorant hypersaline lagoon and saltern ponds (Raina *et al.*, 2010). Determination of the taxonomic composition of these other hypersaline environments could indicate whether selection is occuring for functional capacity or on a taxonomic level if the taxonomic composition between these systems was significantly different but abundance of DMSP lyase genes were high.

The accumulated DMS in Organic Lake suggests conditions in Organic Lake favor the relatively inefficient lysis pathway, where both sulphur and carbon is lost to the organism performing the DMSP lysis, over the more 'thrifty' demethylation pathway. This is particularly pertinent to the *Roseobacter* lineages that can also perform either process. One possibility that has been proposed is that when sulphur is in excess and the organism can easily assimilate alternative sulphur sources, the lysis pathway may be competitive (Johnston *et al.*, 2008). This may be particularly the case in hypersaline systems if higher concentrations of DMSP are being produced as an osmolyte.

4.4.8 Conclusion

Through the use of shotgun metagenomics and size partitioning of samples, we discovered that the Organic Lake system is dominatedby heterotrophic bacteria related to *Psychroflexus*, *Marinobacter* and *Roseovarius* with primary production providedlargely by chlorophyte algae related to *Dunaliella*. Genetic potential for oxidation of fixed carbon by heterotrophic bacteria occurs greatly in excess of carbon fixation, suggesting possible net carbon loss. However, by linking key metabolic processes to the dominant heterotrophic lineages we uncovered processes that were unusually abundant in Organic Lake that may serve to maximize exploitation of limited resources and minimize loss. Recalcitrant polymeric algal material and particulate matter is likely remineralized by *Psychroflexus* in the upper mixed zone and by *Firmicutes* in the deep zone to provide labile substrates for use by other heterotrophic bacteria. The generalist *Marinobacter* and *Roseovarius* lineages were associated with abundant genes involved in rhodopsin-mediated and AAnP photoheterotrophy; the latter of which was more abundant in Organic Lake than any other system surveyed. Potential for chemolithoheterotrophy,

sulphur oxidation and CO oxidation was also high, and along with photo-heterotrophy, may provide a supplementary energy source if organic carbon becomes limiting.

In addition to being able to describe the functional capacities and potential importance of poorly understood microbial processes occurring in the lake (e.g. photoheterotrophy by *Alphaproteobacteria*), we were able to answer targeted questions about the biology of the unusual lake sulphur chemistry. The low potential for dissimilatory sulphur cycling in the deep zone and relatively stable waters, combined with the generation of DMS from DMSP, facilitate the accumulation of a high level of DMS in the lake. It appears *Marinobacter* and *Roseovarius* play a key role in DMS formation by cleaving DMSP generated by upper mixed zone eucaryal algae. The remarkable abundance of DMSP lyase genes suggests DMSP is a significant carbon source in Organic Lake and the cleavage pathway provides a selective advantage under the unique constraints of the Organic Lake environment.

In view of the minimal capacity for biological fixation of carbon and nitrogen, and yet organic richness, including high levels of DMS, in Organic Lake,we evaluated what input the lake may have received throughout its relatively brief $\sim$3,000 year history. The volume of the lake is small ($\sim$6 $\times$ 10^4 m^3), and exogenous input may occur from guano deposits in a small penguin rookery nearby the lake, through giant petrel or skua predation and defecation, and/or by decaying animal carcasses such as elephant seals which can weigh on the order of 1 ton and are present near the lake. It is also possible that during isolation from the ocean, the base of the water column in the marine basin that formed the lake may have acted as a sump for organic material. Phytoplankton blooms and benthic mats tend to make coastal marine basins very productive, and organic matter that sediments out of the surface waters will become trapped in the denser, more saline bottom layers (Bird *et al.*, 1991). Retention of captured organic matter in the lake may also have been facilitated by Organic Lake having become highly saline quickly (Bird *et al.*, 1991). Studies in the future that experimentally determine exogenous input and historical lake dynamics (e.g. stable isotope and biomarker analyses of lake sediment), the role of benthic communities, and metaproteogenomic analyses of interannual community composition and function, will provide improved knowledge of the unusual biogeochemistry of Organic Lake and better enable predictions to be made about how the

lake may be affected by ecosystem changes.

4.5 Acknowledgements

Chapter 5

Ace Lake Viral Genomes

Chapter 6

Conclusions and future work

Some ideas for this section include

Perspective of Antarctic Lake research from wetlab to molecular age. Summary of all molecular work done by our group to present. Summary of the major achievements of my work.

Needed future work for virophages. Since publication of my work, more virophages have been found. Need to isolate and track them over a season. Determine which OLPV type it infects. Verify OLPV infects pyramimonas. Verify OLV reduces infective particles. Make an exclusion experiment to show that the dynamics change with and without the OLV.

References

Abell G. C. J. and Bowman J. P. (2005). Ecological and biogeographic relationships of class Flavobacteria in the Southern Ocean. *FEMS microbiology ecology*, 51(2):265–77.

Abell G. C. J. and Bowman J. P. (2005). Colonization and community dynamics of class Flavobacteria on diatom detritus in experimental mesocosms based on Southern Ocean seawater. *FEMS microbiology ecology*, 53 (3):379–391.

Alekhina I. A., Marie D., Petit J. R., Lukin V. V., Zubkov V. M., and Bulat S. A. (2007). Molecular analysis of bacterial diversity in kerosene-based drilling fluid from the deep ice borehole at Vostok, East Antarctica. *FEMS microbiology ecology*, 59(2):289–299.

Altschul S. F., Gish W., Miller W., Myers E. W., and Lipman D. J. (1990). Basic local alignment search tool. *Journal of molecular biology*, 215:403–410.

Anesio A. M. and Bellas C. M. (2011). Are low temperature habitats hot spots of microbial evolution driven by viruses? *Trends in microbiology*, 19(2):52–57.

Balashov S. P., Imasheva E. S., Boichenko V. a., Antón J., Wang J. M., and Lanyi J. K. (2005). Xanthorhodopsin: a proton pump with a light-harvesting carotenoid antenna. *Science (New York, N.Y.)*, 309(5743): 2061–4.

Bell E. M. and Laybourn-Parry J. (2003). MIXOTROPHY IN THE ANTARCTIC PHYTOFLAGELLATE , PYRAMIMONAS GELIDICOLA (CHLOROPHYTA : PRASINOPHYCEAE). *Journal of Phycology*, 649:644–649.

Bengtsson J., Eriksson K. M., Hartmann M., Wang Z., Shenoy B. D., Grelet G.-A., Abarenkov K., Petri A., Rosenblad M. A., and Nilsson R. H. (2011). Metaxa : a software tool for automated detection and discrimination among ribosomal small subunit mitochondria , and chloroplasts

in metagenomes and environmental sequencing datasets. *Antonie van Leeuwenhoek*, 100:471–475.

Bergh O. i., BØ rsheim K. Y., Bratbak G., and Heldal M. (1989). High abundance of viruses found in aquatic environments. *Nature*, 340:467–468.

Bielewicz S., Bell E. M., Kong W., Friedberg I., Priscu J. C., and Morgan-Kiss R. M. (2011). Protist diversity in a permanently ice-covered Antarctic lake during the polar night transition. *The ISME journal*, 5(9):1559–1564.

Bird M. I., Chivas A. R., Radnell C. J., and Burton H. R. (1991). Sedimentological and stable-isotope evolution of lakes in the Vestfold Hills, Antarctica. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 84:109–130.

Bowman J. P., Mccammon S. A., Lewis T., Skerratt J. H., Brown J. L., Nichols D. S., and Mcmeekin T. A. (1998). Psychroflexus torquis gen. nov., sp. nov., a psychrophilic species from Antarctic sea ice, and reclassification of Flavobacterium gondwanense (Dobson et al. 1993) as Psychroflexus gondwanense gen. nov., comb. nov. *Microbiology*, 144:1601–1609.

Bowman J. P., McCammon S. A., Rea S. M., and McMeekin T. A. (2000). The microbial composition of three limnologically disparate hypersaline Antarctic lakes. *FEMS Microbiology Letters*, 183(1):81–88.

Bowman J. P., Rea S. M., McCammon S. A., and McMeekin T. A. (2000). Diversity and community structure within anoxic sediment from marine salinity meromictic lakes and a coastal meromictic marine basin, Vestfold Hills, Eastern Antarctica. *Environmental Microbiology*, 2(2):227–237.

Bronge C. (2004). Hydrographic and climatic changes influencing the proglacial Druzhby drainage system, Vestfold Hills, Antarctica. *Antarctic Science*, 8(4):379–388.

Budinoff C. R., Loar S. N., LeCleir G. R., Wilhelm S. W., and Buchan A. (2011). A protocol for enumeration of aquatic viruses by epifluorescence microscopy using Anodisc 13 membranes. *BMC microbiology*, 11(1):168.

Bulat S. A., Alekhina I. A., Blot M., Petit J.-R., Angelis M.de, Wagenbach D., Lipenkov V. Y., Vasilyeva L. P., Wloch D. M., Raynaud D., and Lukin V. V. (2004). DNA signature of thermophilic bacteria from the aged accretion ice of Lake Vostok, Antarctica: implications for searching for life in extreme icy environments. *International Journal of Astrobiology*, 3(1):1–12.

Burke C. and Burton H. R. (1988). Photosynthetic bacteria in meromictic lakes and stratified fjords of the Vestfold Hills, Antarctica. *Hydrobiologia*, 165(1):13–23.

Burton H. R. (1981). Chemistry , physics and evolution of Antarctic saline lakes. *Hydrobiologia*, 82(1):339–362.

Caporaso J. G., Kuczynski J., Stombaugh J., Bittinger K., Bushman F. D., Costello E. K., Fierer N., Peña A. G., Goodrich J. K., Gordon J. I., Huttley G. A., Kelley S. T., Knights D., Koenig J. E., Ley R. E., Lozupone C. A., Mcdonald D., Muegge B. D., Pirrung M., Reeder J., Sevinsky J. R., Turnbaugh P. J., Walters W. A., Widmann J., Yatsunenko T., Zaneveld J., and Knight R. (2010). QIIME allows analysis of high- throughput community sequencing data. *Nature Publishing Group*, 7(5):335–336.

Charlson R. J., Lovelock J. E., Andreae M. O., and Warren S. G. (1987). Oceanic phytoplankton, atmospheric sulphur, cloud albedo and climate. *Nature*, 326:655–661.

Christner B. C., Mosley-Thompson E., Thompson L. G., and Reeve J. N. (2001). Isolation of bacteria and 16S rDNAs from Lake Vostok accretion ice. *Environmental Microbiology*, 3(9):570–577.

Clarke K. and RN G. PRIMER V6: User Manual/Tutorial, 2006.

Cottrell M. T. and Kirchman D. L. (2009). Photoheterotrophic microbes in the Arctic Ocean in summer and winter. *Applied and environmental microbiology*, 75(15):4958–4966.

Curson A. R., Rogers R., Todd J. D., Brearley C., and Johnston A. W. (2008). Molecular genetic analysis of a dimethylsulfoniopropionate lyase that liberates the climate-changing gas dimethylsulfide in several marine alpha-proteobacteria and Rhodobacter sphaeroides. *Environmental microbiology*, 10(3):757–767.

Curson A. R., Todd J. D., Sullivan M. J., and Johnston A. W. (2011). Catabolism of dimethylsulphoniopropionate: microorganisms , enzymes and genes. *Nature Reviews Microbiology*, 9:849–859.

Deprez P. P., Franzmann P. D., and Burton H. R. (1986). Determination of reduced sulfur gases in Antarctic lakes and seawater by gas chromatography after solid adsorbent preconcentration. *Journal of Chromatography*, 362:9–21.

Diemer G. S., Kyle J. E., and Stedman K. M. Counting viruses using polycarbonate Track Etch membrane filters as an alternative to Anodisc membrane filters ., 2012. URL `http://www.web.pdx.edu/~kstedman/PCTE_virus_counting_protocol.pdf`.

Dobson S., James S., Franzmann P. D., and Mcmeekin T. A. (1991). A numerical taxonomic study of some pigmented bacteria isolated from Organic Lake, an antarctic hypersaline lake. *Archives of Microbiology*, 156: 56–61.

Fofonoff N. and Millard R. J. (1983). Algorithms for computation of fundamental properties of seawater. *UNESCO technical papers in marine science*, 44.

Franzmann P. D., Burton H. R., and Mcmeekin T. A. (1987). Halomonas subglaciescola, a new species of halotolerant bacteria isolated from Antarctica. *International journal of systematic bacteriology*, 37:27–34.

Franzmann P. D., Deprez P. P., Burton H. R., and Hoff J.van den (1987). Limnology of Organic Lake, Antarctica, a meromictic lake that contains high concentrations of dimethyl sulfide. *Marine and Freshwater Research*, 38(3):409–417.

Franzmann P. D., Stackebrandt E., Sanderson K., Volkman J. K., Cameron D., Stevenson P., and Mcmeekin T. A. (1988). Halorubrum lacusprofundi sp. nov., a halophilic bacterium idolated from Deep Lake, Antarctica. *Systematic & Applied Microbiology*, 11:20–27.

Friedrich C. G., Bardischewsky F., Rother D., Quentmeier A., and Fischer J. (2005). Prokaryotic sulfur oxidation. *Current opinion in microbiology*, 8(3):253–259.

Fuhrman J. A., Schwalbach M. S., and Stingl U. (2008). Proteorhodopsins: an array of physiological roles? *Nature reviews. Microbiology*, 6:488–494.

Gibson J. A. (1999). The meromictic lakes and stratified marine basins of the Vestfold Hills, East Antarctica. *Antarctic Science*, 11(2):175–192.

Gibson J. A., Ferris J. M., and Burton H. R. (1990). Temperature density, temperature conductivity and conductivity-density relationships for marine-derived saline lake waters. *ANARE research notes*, 78.

Gibson J. A., Garrick R., Franzmann P. D., Deprez P. P., and Burton H. R. (1991). Reduced sulfur gases in saline lakes of the Vestfold Hills, Antarctica. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 84(1-4):131–140.

Gibson J. A., Qiang X. L., Franzmann P. D., Garrick R. C., and Burton H. R. (1994). Volatile fatty and dissolved free amino acids in Organic Lake, Vestfold Hilss, East Antarctica. *Polar Biology*, 14:545–550.

Glatz R., Lepp P., Ward B. B., and Francis C. (2006). Planktonic microbial community composition across steep physical/chemical gradients in permanently ice-covered Lake Bonney, Antarctica. *Geobiology*, 4:53–67.

Gómez-Consarnau L., González J. M., Coll-Lladó M., Gourdon P., Pascher T., Neutze R., Pedrós-Alió C., and Pinhassi J. (2007). Light stimulates growth of proteorhodopsin- containing marine Flavobacteria. 445:210–213.

Gómez-Consarnau L., Akram N., Lindell K., Pedersen A., Neutze R., Milton D. L., González J. M., and Pinhassi J. (2010). Proteorhodopsin phototrophy promotes survival of marine bacteria during starvation. *PLoS Biology*, 8(4):2–11.

Gordon D., Priscu J. C., and Giovanonni S. J. (2000). Origin and Phylogeny of Microbes Living in Permanent Antarctic Lake Ice. *Microbial ecology*, 39(3):197–202.

Green W. J., Angle M. P., and Chave K. E. (1988). The geochemistry of Antarctic streams and their role in the evolution of four lakes of the McMurdo Dry Valleys. *Geochimica et Cosmochimica Acta*, 52(5):1265–1274.

Hahsler M., Hornik K., and Buchta C. (2007). Getting things in order: an introduction to the R package seriation. *Journal of Statistical Software*, 25(3):1–34.

Hodgson D. A. Antarctic Lakes. In Bengtsson L., Herschy R. W., and Fairbridge R. W., editors, *Encyclopedia of Lakes and Reservoirs*, pages 26–31. Springer, Dordrecht, 2012.

Howard E. C., Henriksen J. R., Buchan A., Reisch C. R., Bürgmann H., Welsh R., Ye W., González J. M., Mace K., Joye S. B., Kiene R. P., Whitman W. B., and Moran M. A. (2006). Bacterial taxa that limit sulfur flux from the ocean. *Science (New York, N.Y.)*, 314(5799):649–652.

Howard E. C., Sun S., Biers E. J., and Moran M. A. (2008). Abundant and diverse bacteria involved in DMSP degradation in marine surface waters. *Environmental microbiology*, 10(9):2397–2410.

Huang Y., Niu B., Gao Y., Fu L., and Li W. (2010). CD-HIT Suite : a web server for clustering and comparing biological sequences. *Bioinformatics*, 26:260–262.

Johnston A. W., Todd J. D., Sun L., Nikolaidou-Katsaraidou N., Curson A. R., and Rogers R. (2008). Molecular diversity of bacterial production of the climate-changing gas, dimethyl sulphide, a molecule that impinges on local and global symbioses. *Journal of experimental botany*, 59(5): 1059–1067.

Johnstone G., Brown D., and Lugg D. (1973). The biology of the Vestfold Hills, Antarctica. *ANARE Scientific Reports*, 123:1–60.

Karr E. A., Sattley W. M., Jung D. O., Madigan M. T., and Achenbach L. A. (2003). Remarkable Diversity of Phototrophic Purple Bacteria in a Permanently Frozen Antarctic Lake. 8(69):4910–4914.

Karr E. A., Sattley W. M., Rice M. R., Jung D. O., Madigan M. T., and Achenbach L. A. (2005). Diversity and Distribution of Sulfate-Reducing Bacteria in Permanently Frozen Lake Fryxell , McMurdo Dry Valleys , Antarctica. *Applied and Environmental Microbiology*, 71:6353–6359.

Karr E. A., Ng J. M., Belchik S. M., Matthew W. M., Madigan M. T., Achenbach L. A., and Sattley W. M. (2006). Biodiversity of Methanogenic and Other Archaea in the Permanently Frozen Lake Fryxell , Antarctica. 72(2):1663–1666.

Kepner R. L., Wharton R. A., and Suttle C. A. (1998). Viruses in Antarctic lakes. *Limnology and oceanography*, 43(7):1754–1761.

Kurosawa N., Sato S., Kawarabayasi Y., Imura S., and Naganuma T. (2010). Archaeal and bacterial community structures in the anoxic sediment of Antarctic meromictic lake Nurume-Ike. *Polar Science*, 4(2):421–429.

Labrenz M., Collins M. D., Lawson P. A., Tindall B. J., Schumann P., and Hirsch P. (1999). Roseovarius tolerans gen. nov., sp. nov., a budding bacterium with variable bacteriochlorophyll a production from hypersaline Ekho Lake. *International Journal of Systematic Bacteriology*, 49:137–147.

Lauro F. M., Demaere M. Z., Yau S., Brown M. V., Ng C., Wilkins D., Raftery M. J., Gibson J. A., Andrews-Pfannkoch C., Lewis M., Hoffman J. M., Thomas T., and Cavicchioli R. (2011). An integrative study of a meromictic lake ecosystem in Antarctica. *The ISME journal*, 5:879–895.

Law P. (1959). The Vestfold Hills. *ANARE Reports*, 1:1–50.

Laybourn-Parry J. The microbial loop in Antarctic lakes. In Howard-Willsiams C., Lyons W., and Hawes I., editors, *Ecosystem Dynamics in a Antarctic Ice-Free Landscapes*, pages 231–240. Rotterdam, 1997.

Laybourn-Parry J. and Pearce D. A. (2007). The biodiversity and ecology of Antarctic lakes: models for evolution. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 362:2273–2289.

Laybourn-Parry J., Hofer J. S., and Sommaruga R. (2001). Viruses in the plankton of freshwater and saline Antarctic lakes. *Freshwater Biology*, 46 (9):1279–1287.

Laybourn-Parry J., Marshall W. A., and Marchant H. J. (2005). Flagellate nutritional versatility as a key to survival in two contrasting Antarctic saline lakes. *Freshwater Biology*, 50:830–838.

López-Bueno A., Tamames J., Velázquez D., Moya A., Quesada A., and Alcamí A. (2009). High diversity of the viral community from an Antarctic lake. *Science*, 326:858–861.

Lovelock J. E. and Maggs R. (1972). Atmospheric dimethyl sulphide and the natural sulphur cycle. *Nature*, 237:452–453.

Madan N. J., Marshall W. A., and Laybourn-Parry J. (2005). Virus and microbial loop dynamics over an annual cycle in three contrasting Antarctic lakes. *Freshwater B*, 50:1291–1300.

Man D., Wang W., Sabehi G., Aravind L., Post A. F., Massana R., Spudich E. N., Spudich J. L., and Béjà O. (2003). Diversification and spectral tuning in marine proteorhodopsins. *The EMBO Journal*, 22(8):1725–1731.

Matsuzaki M., Kubota K., Satoh T., Kunugi M., Ban S., and Imura S. (2006). Dimethyl sulfoxide-respiring bacteria in Suribati Ike, a hypersaline lake, in Antarctica and the marine environment. *Polar Bioscience*, 20:73–81.

Mikucki J. A. and Priscu J. C. (2007). Bacterial diversity associated with Blood Falls, a subglacial outflow from the Taylor Glacier, Antarctica. *Applied and environmental microbiology*, 73(12):4029–39.

Mikucki J. A., Pearson A., Johnston D. T., Turchyn A. V., Farquhar J., Schrag D. P., Anbar A. D., Priscu J. C., and Lee P. A. (2009). A contemporary microbially maintained subglacial ferrous "ocean". *Science (New York, N.Y.)*, 324(5925):397–400.

Moran M. A. and Miller W. L. (2007). Resourceful heterotrophs make the most of light in the coastal ocean. *Nature reviews. Microbiology*, 5(10): 792–800.

Moran M. A., Reisch C. R., Kiene R. P., and Whitman W. B. (2012). Genomic Insights into Bacterial DMSP Transformations. *Annual Review of Marine Science*, 4(1):523–542.

Mosier A. C., Murray A. E., and Fritsen C. H. (2007). Microbiota within the perennial ice cover of Lake Vida, Antarctica. *FEMS microbiology ecology*, 59(2):274–288.

Ng C. *A metaproteomic analysis of microbial communities in Ace Lake, Antarctic*. PhD thesis, University of New South Wales, 2010.

Ng C., DeMaere M. Z., Williams T. J., Lauro F. M., Raftery M., Gibson J. A., Andrews-Pfannkoch C., Lewis M., Hoffman J. M., Thomas T., and Cavicchioli R. (2010). Metaproteogenomic analysis of a dominant green sulfur bacterium from Ace Lake, Antarctica. *The ISME journal*, 4(8): 1002–19.

Noble R. T. and Fuhrman J. A. (1998). Use of SYBR Green I for rapid epifluorescence counts of marine viruses and bacteria. *Aquatic Microbial Ecology*, 14:113–118.

Noguchi H., Park J., and Takagi T. (2006). MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. *Nucleic acids research*, 34(19):5623–5630.

Olson J., Steppe T., Litaker R., and Paerl H. (1998). N2-Fixing Microbial Consortia Associated with the Ice Cover of Lake Bonney, Antarctica. *Microbial ecology*, 36(3):231–238.

Patel A., Noble R. T., Steele J. A., Schwalbach M. S., Hewson I., and Fuhrman J. A. (2007). Virus and prokaryote enumeration from planktonic aquatic environments by epifluorescence microscopy with SYBR Green I. *Nature protocols*, 2(2):269–276.

Pearce D. A. (2003). Bacterioplankton community structure in a maritime antarctic oligotrophic lake during a period of holomixis, as determined by denaturing gradient gel electrophoresis (DGGE) and fluorescence in situ hybridization (FISH). *Microbial Ecology*, 46(1):92–105.

Pearce D. A. (2005). The structure and stability of the bacterioplankton community in Antarctic freshwater lakes, subject to extremely rapid environmental change. *FEMS microbiology ecology*, 53(1):61–72.

Pearce D. A., Gast C. J., Lawley B., and Ellis-Evans J. C. (2003). Bacterioplankton community diversity in a maritime Antarctic lake, determined by culture-dependent and culture-independent techniques. *FEMS microbiology ecology*, 45(1):59–70.

Pearce D. A., Gast C. J.van der, Woodward K., and Newsham K. K. (2005). Significant changes in the bacterioplankton community structure of a maritime Antarctic freshwater lake following nutrient enrichment. *Microbiology (Reading, England)*, 151:3237–3248.

Pickard J., Adamson D. A., and Heath C. W. (1986). The evolution of Watts Lake, Vestfold Hills, East Antarctica, from marine inlet to freshwater lake. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 53:271–288.

Priscu J. C. (1999). Geomicrobiology of Subglacial Ice Above Lake Vostok, Antarctica. *Science*, 286(5447):2141–2144.

Proctor L. M. and Fuhrman J. A. (1990). Viral mortality of marine bacteria and cyanobacteria. *Nature*, 343:60–62.

Purdy K., Nedwell D., and Embley T. (2003). Analysis of the sulfate-reducing bacterial and methanogenic archaeal populations in contrasting Antarctic sediments. *Applied and environmental . . .*, 69(6):3181–3191.

Raina J.-B., Dinsdale E. A., Willis B. L., and Bourne D. G. (2010). Do the organic sulfur compounds DMSP and DMS drive coral microbial associations? *Trends in microbiology*, 18(3):101–108.

Reisch C. R., Moran M. A., and Whitman W. B. (2011). Bacterial catabolism of dimethylsulfoniopropionate (DMSP). *Frontiers in microbiology*, 2(August):1–12.

Roberts N. and Burton H. R. (1993). Sampling volatile organics from a meromictic Antarctic lake. *Polar Biology*, 13:359– 361.

Roberts N., Burton H. R., and Pitson G. (1993). Volatile organic compounds from Organic Lake, an Antarctic, hypersaline, meromictic lake. *Antarctic Science*, 5(4):361–366.

Rusch D. B., Halpern A. L., Sutton G., Heidelberg K. B., Williamson S. J., Yooseph S., Wu D., Eisen J. A., Hoffman J. M., Remington K., Beeson K. Y., Tran B., Smith H., Baden-Tillson H., Stewart C., Thorpe J., Freeman J., Andrews-Pfannkoch C., Venter J. E., Li K., Kravitz S., Heidelberg J. F., Utterback T., Rogers Y.-H., Falcón L. I., Souza V., Bonilla-Rosso G., Eguiarte L. E., Karl D. M., Sathyendranath S., Platt T., Bermingham E., Gallardo V., Tamayo-Castillo G., Ferrari M. R., Strausberg R. L., Nealson K., Friedman R., Frazier M., and Venter J. C. (2007). The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS biology*, 5(3):e77.

Säwström C., Anesio A. M., Granéli W., and Laybourn-Parry J. (2007). Seasonal viral loop dynamics in two large ultraoligotrophic Antarctic freshwater lakes. *Microbial ecology*, 53:1–11.

Schäfer H., Myronova N., and Boden R. (2010). Microbial degradation of dimethylsulphide and related C1-sulphur compounds: organisms and pathways controlling fluxes of sulphur in the biosphere. *Journal of experimental botany*, 61(2):315–334.

Schiaffino M. R., Unrein F., Gasol J. M., Farias M. E., Estevez C., Balagué V., and Izaguirre I. (2009). Comparative analysis of bacterioplankton assemblages from maritime Antarctic freshwater lakes with contrasting trophic status. *Polar Biology*, 32(6):923–936.

Sharma A. K., Zhaxybayeva O., Papke R. T., and Doolittle W. F. (2008). Actinorhodopsins: proteorhodopsin-like gene sequences found predominantly in non-marine environments. *Environmental microbiology*, 10(4): 1039–1056.

Siegert M. J., Ellis-Evans J. C., Tranter M., Mayer C., Petit J.-R., Salamatin A., and Priscu J. C. (2001). Physical, chemical and biological processes

in Lake Vostok and other Antarctic subglacial lakes. *Nature*, 414(6864): 603–609.

Siegert M. J., Carter S., Tabacco I., Popov S., and Blankenship D. D. (2005). A revised inventory of Antarctic subglacial lakes. *Antarctic Science*, 17 (03):453–460.

Tamura K., Peterson D., Peterson N., Stecher G., Nei M., and Kumar S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution*, 28(10):2731–2739.

Todd J. D., Rogers R., Li Y. G., Wexler M., Bond P. L., Sun L., Curson A. R., Malin G., Steinke M., and Johnston A. W. (2007). Structural and regulatory genes required to make the gas dimethyl sulfide in bacteria. *Science (New York, N.Y.)*, 315(5812):666–669.

Todd J. D., Curson A. R., Dupont C. L., Nicholson P., and Johnston A. W. (2009). The dddP gene, encoding a novel enzyme that converts dimethylsulfoniopropionate into dimethyl sulfide, is widespread in ocean metagenomes and marine bacteria and also occurs in some Ascomycete fungi. *Environmental microbiology*, 11(6):1376–1385.

Todd J. D., Curson A. R., Nikolaidou-Katsaraidou N., Brearley C. A., Watmough N. J., Chan Y., Page P. C., Sun L., and Johnston A. W. (2010). Molecular dissection of bacterial acrylate catabolism - unexpected links with dimethylsulfoniopropionate catabolism and dimethyl sulfide production. *Environmental microbiology*, 12(2):327–343.

Todd J. D., Curson A. R., Kirkwood M., Sullivan M. J., Green R. T., and Johnston A. W. (2011). DddQ, a novel, cupin-containing, dimethylsulfoniopropionate lyase in marine roseobacters and in uncultured marine bacteria. *Environmental microbiology*, 13(2):427–438.

Torrice M. Viral ecology hit by filter shortage, 2009. URL `http://news.sciencemag.org/scienceinsider/2009/10/viral-ecology-r.html`.

Unrein F., Izaguirre I., Massana R., Balagué V., and Gasol J. M. (2005). Nanoplankton assemblages in maritime Antarctic lakes : characterisation and molecular fingerprinting comparison. *Aquatic Microbial Ecology*, 40: 269–282.

Varin T., Lovejoy C., Jungblut A. D., Vincent W. F., and Corbeil J. (2012). Metagenomic analysis of stress genes in microbial mat communities from Antarctica and the High Arctic. *Applied and environmental microbiology*, 78(2):549–59.

Venter J. C., Remington K., Heidelberg J. F., Halpern A. L., Rusch D., Eisen J. A., Wu D., Paulsen I., Nelson K. E., Nelson W., Fouts D. E., Levy S., Knap A. H., Lomas M. W., Nealson K., White O., Peterson J., Hoffman J., Parsons R., Baden-Tillson H., Pfannkoch C., Rogers Y.-H., and Smith H. O. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science (New York, N.Y.)*, 304(5667):66–74.

Villaescusa J. A., Casamayor E. O., Rochera C., Velázquez D., Álvaro C., Quesada A., and Camacho A. (2010). A close link between bacterial community composition and environmental heterogeneity in maritime Antarctic lakes. *International . . .*, (126):67–77.

Voytek M. A., Priscu J. C., and Ward B. B. (1999). The distribution and relative abundance of ammonia-oxidizing bacteria in lakes of the McMurdo Dry Valley, Antarctica. *Hydrobiologia*, 401:113–130.

Wang Q., Garrity G. M., Tiedje J. M., and Cole J. R. (2007). Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and environmental microbiology*, 73(16):5261–5267.

Ward B. B. and Priscu J. C. (1997). Detection and characterization of denitrifying bacteria from a permanently ice-covered Antarctic lake. *Hydrobiologia*, (347):57–68.

Ward B. B., Granger J., Maldonado M., Casciotti K., Harris S., and Wells M. (2005). Denitrification in the hypolimnion of permanently ice-covered Lake Bonney, Antarctica. *Aquatic . . .*, 38(2000):295–307.

Wilkins D., Lauro F. M., Williams T. J., Demaere M. Z., Brown M. V., Hoffman J. M., Andrews-Pfannkoch C., Mcquaid J. B., Riddle M. J., Rintoul S. R., and Cavicchioli R. (2012). Biogeographic partitioning of Southern Ocean microorganisms revealed by metagenomics. *Environmental microbiology*.

Yamamoto M. and Takai K. (2011). Sulfur metabolisms in epsilon- and gamma-Proteobacteria in deep-sea hydrothermal fields. *Frontiers in microbiology*, 2:1–8.

Yau S., Lauro F. M., Demaere M. Z., Brown M. V., Thomas T., Raftery M. J., Andrews-Pfannkoch C., Lewis M., Hoffman J. M., Gibson J. A., and Cavicchioli R. (2011). Virophage control of antarctic algal host virus dynamics. *Proceedings of the National Academy of Sciences of the United States of America*, 108:6163–6168.

Zwartz D., Bird M., Stone J., and Lambeck K. (1998). Holocene sea-level change and ice-sheet history in the Vestfold Hills, East Antarctica. *Earth and Planetary Science Letters*, 155(1-2):131–145.

Chapter 7

Appendices